

# Building forecasting model for multi varied products using R shiny at Hexagon Geo systems India

Vishal Raina <sup>1</sup>

<sup>1</sup> Symbiosis Institute of Operations Management, Nashik, Symbiosis International (Deemed University), Pune, India

## ABSTRACT

Vishal worked as an intern at Hexagon Geosystems, India wherein he was assigned a role of operations intern and specified a task to improve forecasting capabilities for a wide portfolio of products. The as-is forecasting techniques were concentrated around straight line and moving average models that give promising results when the product lies in a smooth category. However, the basic underlying problem that Vishal had to face was that 80% of the products fell under the intermittent, Erratic and Lumpy categories which are supposed to have unreliable forecasts when forecasted using the above-mentioned models. Products are categorized under 4 major clusters that significantly determine their forecasting capabilities based on Average Demand Interval and Coefficient of Variation. So, Vishal kicked off his project by categorizing these products and with a detailed exploratory analysis over the data set. For all but the smooth demand profile, forecast accuracy is not a reliable performance metric. It lacks contextual information and, in the end, leads you to miss the big picture. This induces overstock situations or, on the contrary, poor service level, both situations you want to avoid. This is why you should take some time to understand your products' various demand patterns, step back, and adjust your expectations.

It was crucial to understand products' demand patterns and optimize time and resource allocation for forecasting more than 500 products at a single click of a button. Vishal was given the task of developing a robust forecasting model that would be capable of forecasting massive number of products at minimum time allocation.

The case briefs upon various forecasting models used and deals with the process and techniques of building a forecasting model for massive number of products using interactive capabilities in R Shiny and Excel as a basic data repository.

*Article Received: 10 August 2020, Revised: 25 October 2020, Accepted: 18 November 2020*

## Introduction

Hexagon is a digital solutions company leading the creation of autonomous connected ecosystems (ACE) in the markets it serves. Hexagon's Smart X solutions combine sensors and software to move customers beyond the data impasse of IoT, so that they are able to leverage the vast potential of data being generated by connected things. Through embedded artificial intelligence (AI), edge-cloud orchestration, mobility, and advanced visualisation capabilities, customers are able to move well beyond automation and introduce autonomous processes and decision-making into their workflows to drive further improvements in quality and productivity.

Hexagon delivers digital solutions that improve productivity and quality across manufacturing, infrastructure, safety and mobility applications. Hexagon aspires to be the leader in creating autonomous connected ecosystems for the industries we serve.

ECOSYSTEMS SERVED: • Production – manufacturing, power & energy, mining and agriculture.

• Autonomous – autonomous transport and unmanned aerial/ground/marine vehicles.

• Urban – construction, infrastructure management and public safety.

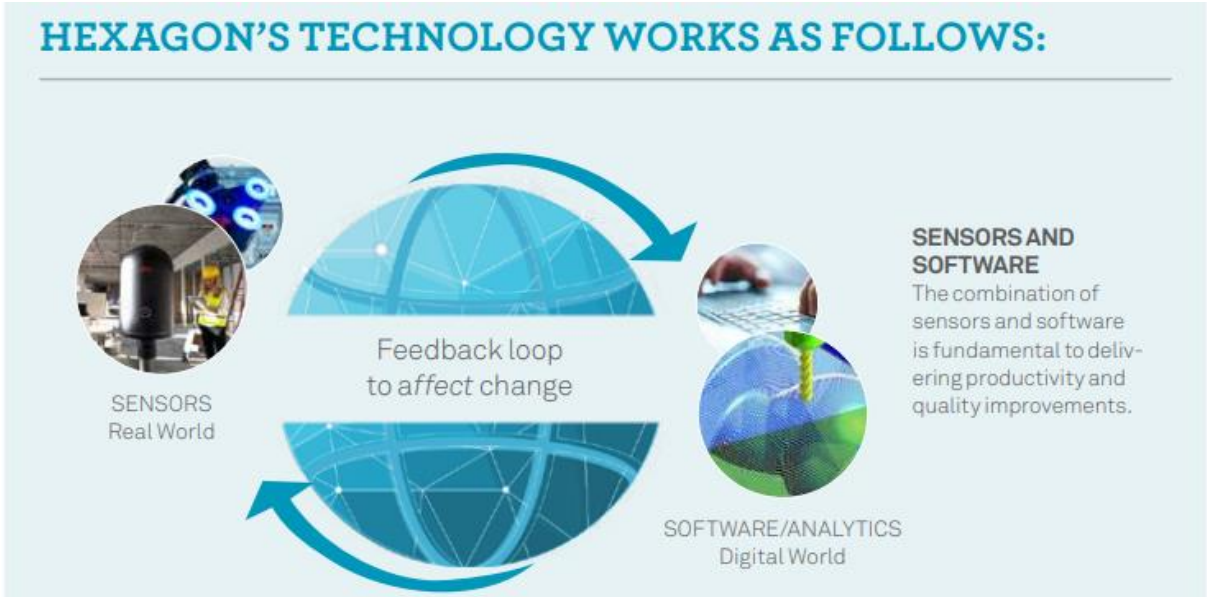


Exhibit 1

**Geospatial Enterprise Solutions (GES)** includes a world-leading portfolio of reality-capture sensors – from laser scanners, airborne cameras and UAVs (unmanned aerial vehicles) to monitoring equipment, mobile mapping technologies and precise positioning. The sensors

are complemented by software for the creation of 3D maps and models which are used for decision-making in a range of software applications, covering areas such as surveying, construction, public safety and agriculture.

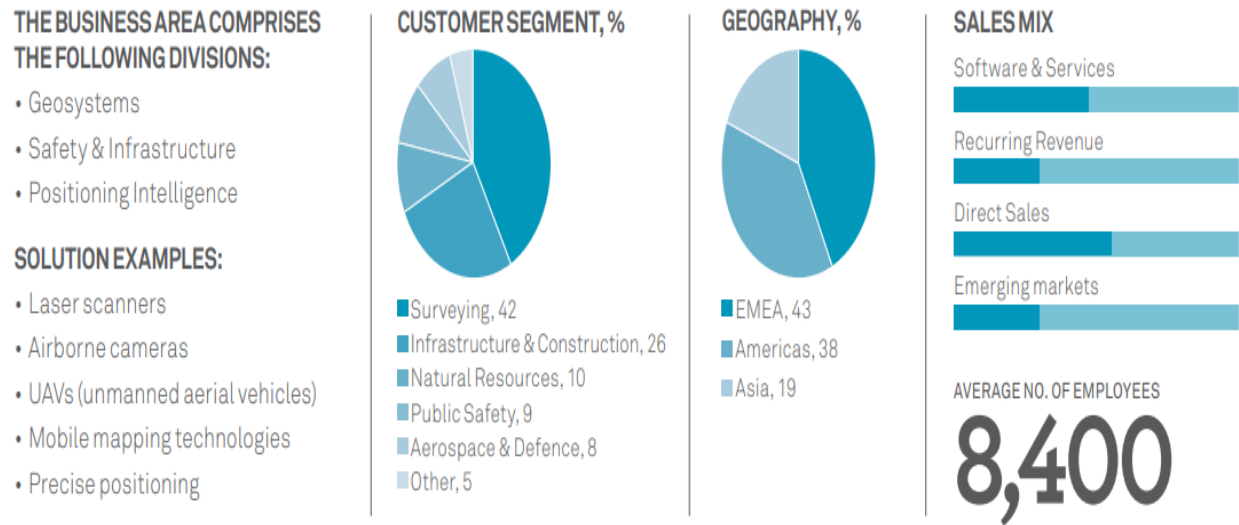


Exhibit 2

**Industrial Enterprise Solutions (IES)** includes a world-leading portfolio of metrology systems that incorporate the latest in sensor technology for fast and accurate measurements. These solutions include technologies such as coordinate measurement machines (CMM) and laser trackers and scanners – which optimise design, processes and throughput in manufacturing facilities. It also includes software for CAD (computer-aided

design), CAM (computer-aided manufacturing) and CAE (computer-aided engineering). Solutions within this segment optimise design and processes, improve productivity in process facilities and create and leverage asset management information critical to the planning, construction and operation of plants and process facilities.

#### THE BUSINESS AREA COMPRISES THE FOLLOWING DIVISIONS:

- Manufacturing Intelligence
- PPM

#### SOLUTION EXAMPLES:

- CMMs (coordinate measuring machines)
- Optical and portable scanners
- Industrial metrology software
- CAD/CAM/CAE (computer-aided design/manufacturing/engineering) software
- Life cycle engineering software

#### CUSTOMER SEGMENT, %



- Electronics & Manufacturing, 32
- Power & Energy, 24
- Automotive, 22
- Aerospace & Defence, 14
- Other, 8

#### GEOGRAPHY, %



- Asia, 39
- EMEA, 32
- Americas, 29

#### SALES MIX

Software & Services



Recurring Revenue



Direct Sales



Emerging markets



AVERAGE NO. OF EMPLOYEES

10,700

### Exhibit 3

#### India Scale of Operations and Projects

Hexagon India aspires to play a leading role in the efforts to solve the challenges faced by our nation by providing ground-breaking information technologies that make a positive and lasting impact. Hexagon is committed to leveraging digital solutions to fundamentally change the way business is done. With the intent to create a connected ecosystem, Hexagon India has been incessantly widening its horizon, engaging with some of the top development projects in India.

The major headways include:

**Dam Safety and Monitoring:** The dams in North East are part of run of the river. These concrete gravity and rock filled dams have been commissioned more than 20 years ago. As the concerned states are in the area of seismic activity, these dams pose threat to the civilization there. Moreover, when the spillway is opened, flow from some of these dams effects the downstream areas, like causing floods. Additionally, the geo-technical equipment installed during the erection of these dams are non-functional and no analysis information of structural characteristics of these dams are available. To combat these issues, Hexagon developed a Dam Safety Monitoring System, a first of its kind in India. The system combined multiple technologies with geodetic technologies to become an effective monitoring application. Under the Government's DRIP (Dam Rehabilitation and Implementation Program), Hexagon became a front runner for a dynamic dam safety monitoring solution.

**Land Records Management:** This is done through Hexagon's Digital India Land Records Modernization program. Hexagon conducted PAN India training as a capacity building initiative of Department of Land Resources. It is closely associated with the Department for Land Records Management, Survey/Resurvey, Mapping & Updating Land Parcel and Management of Land Records.

**Utility detection:** Technological solution from Hexagon is enabling detection of underground utility facilities.

**Mine management:** Integrates planning, operations and safety solutions for the mining industry. The technology integrates the data processes from these solutions and the customer is better informed to make an intelligent decision. Currently, it is catering to 5 mines of Coal India.

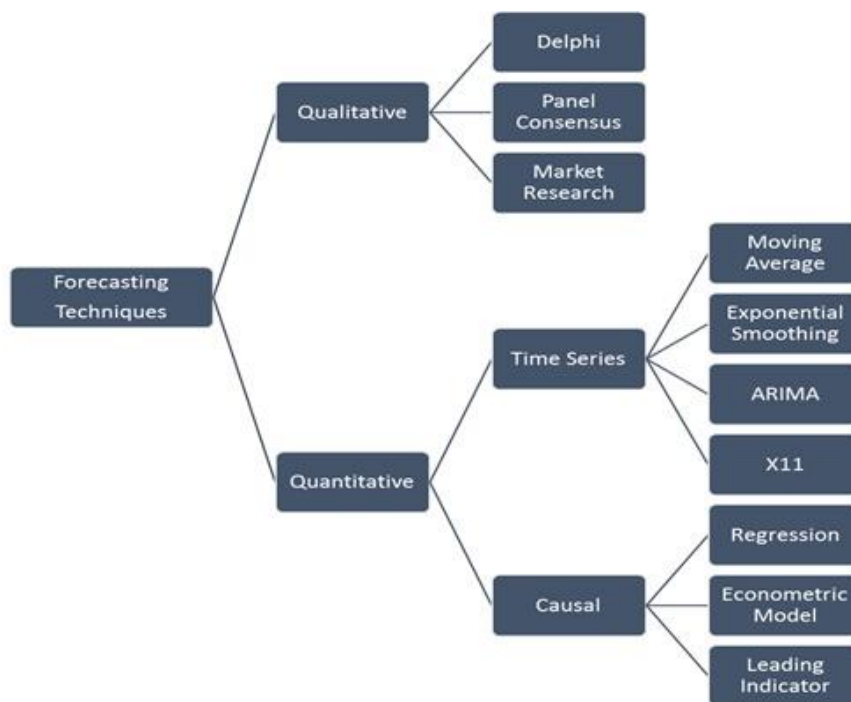
**Hydrography solution:** Hexagon has provided a solution to the Survey of India, for mapping assets and resources for hydrography, both in 2D and 3D. This is done through Hexagon's IMAGINE Suite including photogrammetry module.

**'Mahabhumi' project of Maharashtra Remote Sensing Application Centre:** This project by Hexagon's Geospatial Division incorporates large scale mapping of the whole state. Hexagon's Safety and Infrastructure segment has an extensive portfolio of end to end solutions and domain expertise for orchestration of vital public services that are uniquely positioned to realize safe city strategies. Its Safe City Framework includes mission-critical, safe city-ready components that advance a city's capabilities.

**Forecasting Challenge at Hexagon:** For a diversified range of operating sectors as evident from the projects mentioned above it is crucial for an organisation to have a detailed insight into their supply chain capabilities to prevent any disruptions that may be caused due to natural, economic or political constraints. Moreover, these projects don't render any sort of trend or seasonal characteristics as a majority of them are influenced by the geographical and political arms which don't come under an impulse buying segment. Indian scale of operations is fulfilled by a major distribution centre located in Singapore. This incorporated the concept of lead time for wide range of products and its impact on the safety stock.

### Current Forecasting at Hexagon:

The as-is forecasting techniques were concentrated around straight line and moving average models that give promising results when the product lies in a smooth category. However due to high demand variability these models came along with lower accuracy and higher RMSE (Root Mean Squared Error) and MAPE (Mean Percentage Absolute Error). Alternative Techniques include using Causal Forecasting model, Hierarchical clustering models, Batch forecasting and Crost Forecasting model. These are applicable for a wide portfolio of products and are widely used due to their intermittence resolving capabilities.



**Exhibit 4**

### Steps involved in Forecasting

#### 1: Define the problem.

Frequently this is the most troublesome piece of gauging. Characterizing the issue cautiously requires a comprehension of the manner in which the gauges will be utilized, who requires the estimates, and how the gauging capacity fits inside the association requiring the figures. A forecaster needs to invest energy conversing with each and every individual who will be associated with gathering information, looking after databases, and utilizing the conjectures for future arranging.

#### 2: Collect the available information.

There are consistently in any event two sorts of data required: (a) measurable information, and (b) the amassed mastery of the individuals who gather the information and utilize the conjectures. Frequently, it will be hard to acquire enough chronicled information to have the option to fit a decent factual model. All things considered; the critical gauging techniques can be utilized. Once in a while, old information will be less helpful because of basic changes in the framework being conjecture; at that point we may decide to utilize just the latest information. In any case, recollect that great factual models will deal with



transformative changes in the framework; don't discard great information pointlessly.

### **3: Perform Descriptive and exploratory analysis.**

Continuously start by charting the information. Are there reliable examples? Is there a huge pattern? Is irregularity significant? Is there proof of the nearness of business cycles? Are there any anomalies in the information that should be clarified by those with master information? How solid are the connections among the factors accessible for investigation? Different instruments have been created in R to help with this examination.

### **4: Choose data relevant fitting models.**

The best model to utilize relies upon the accessibility of authentic information, the quality of connections between the figure variable and any logical factors, and the manner by which the estimates are to be utilized. It isn't unexpected to look at a few likely models. Each model is itself a fake develop that depends on a series of expectations (unequivocal and certain) and as a rule includes at least one boundary which must be evaluated utilizing the known chronicled information.

### **Step 5: Evaluate and use the model.**

When a model has been chosen and its boundaries evaluated, the model is utilized to make figures. The exhibition of the model must be appropriately assessed after the information for the figure time frame have opened up. Various techniques have been created to help in evaluating the precision of conjectures. There are likewise authoritative issues in utilizing and following up on the conjectures.

## **Description about Forecasting Techniques**

- **Simple Moving Average:** The simplest form of a moving average, known as a simple moving average (SMA), is calculated by taking the arithmetic mean of a given set of values. In other words, a set of numbers—or sales values in the case of SKUs or product level—are added together and then divided by the number of sales values in the set.
- **Exponential Moving Average:** The exponential moving average is a type of moving average that gives more weight to recent sales in an attempt to make it more responsive to new information. To

calculate an EMA, you must first compute the simple moving average (SMA) over a particular time period. Next, you must calculate the multiplier for weighting the EMA (referred to as the "smoothing factor"). Then you use the smoothing factor combined with the previous EMA to arrive at the current value. The EMA thus gives a higher weighting to recent prices, while the SMA assigns equal weighting to all values.

- **Holt-Winters Forecasting:** Holt-Winters forecasting is a way to model and predict the behavior of a sequence of values over time—a time series. Holt-Winters is one of the most popular forecasting techniques for time series. It's decades old, but it's still ubiquitous in many applications, including monitoring, where it's used for purposes such as anomaly detection and capacity planning. Holt-Winters is a model of time series behavior. Forecasting always requires a model, and Holt-Winters is a way to model three aspects of the time series: a typical value (average), a slope (trend) over time, and a cyclical repeating pattern (seasonality). Holt-Winters uses exponential smoothing to encode lots of values from the past and use them to predict "typical" values for the present and future.

The three aspects of the time series behavior—value, trend, and seasonality—are expressed as three types of exponential smoothing, so Holt-Winters is called triple exponential smoothing. The model predicts a current or future value by computing the combined effects of these three influences. The model requires several parameters: one for each smoothing ( $\alpha$ ,  $\beta$ ,  $\gamma$ ), the length of a season, and the number of periods in a season.

Seasonality can be confusing. A season is a fixed length of time that contains the full repetition. You might think your data repeats daily (there's a peak at 2pm every day), but if the weekend has different behavior (there's no peak at 2pm on Sunday) then your season is really a week, not a day. Within the season, there are periods, which is the granularity of prediction. If you want to model a value for every hour of every day within a week,

your season is 168 hours long and your period is 1 hour.

- **ARIMA Forecasting:** ARIMA, short for 'Auto Regressive Integrated Moving Average' is actually a class of models that 'explains' a given time series based on its own past values, that is, its own lags and the lagged forecast errors, so that equation can be used to forecast future values. Any 'non-seasonal' time series that exhibits patterns and is not a random white noise can be modeled with ARIMA models. An ARIMA model is characterized by 3 terms: p, d, q where,

p is the order of the AR term

q is the order of the MA term

d is the number of differencing required to make the time series stationary

If a time series, has seasonal patterns, then you need to add seasonal terms and it becomes SARIMA, short for 'Seasonal ARIMA'. The first step to build an ARIMA model is to [make the time series stationary](#). Because, term 'Auto Regressive' in ARIMA means it is a [linear regression model](#) that uses its own lags as predictors. Linear regression models, as you know, work best when the predictors are not correlated and are independent of each other. So how to make a series stationary?

The most common approach is to difference it. That is, subtract the previous value from the current value. Sometimes, depending on the complexity of the series, more than one

differencing may be needed. The value of d, therefore, is the minimum number of differencing needed to make the series stationary. And if the time series is already stationary, then  $d = 0$ .

'p' is the order of the 'Auto Regressive' (AR) term. It refers to the number of lags of Y to be used as predictors. And 'q' is the order of the 'Moving Average' (MA) term. It refers to the number of lagged forecast errors that should go into the ARIMA Model.

**ARIMA model in words:** Predicted  $Y_t$  = Constant + Linear combination Lags of Y (upto p lags) + Linear Combination of Lagged forecast errors (upto q lags)

### Task Ahead

A robust forecasting tool can help to better forecast the product sales and then consolidate the reports with existing inventory reports to calculate an optimal level of safety stock and a suitable product specific reorder point.

Vishal decides to develop a forecasting tool and starts digging into what data shall be needed to work on the tool. What all information needs to be gathered to work on the model. Vishal makes a list, time duration for time series data to capture relevant characteristics and run statistical confirmatory analysis.

For the tool to work in an efficient way, Vishal decided to collect historical sales data of all the products, further he arranges the data in a tabular form in a csv file format which can be read by the Shiny code.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	
1	Date	Product 1	Product 2	Product 3	Product 4	Product 5	Product 6	Product 7	Product 8	Product 9	Product 10	Product 11	Product 12	Product 13	Product 14	Product 15	Product 16	Product 17	Product 18	Pi
2	01-01-2018	42	0	8	0	104	0	0	0	0	279	0	0	0	0	45	0	0	0	0
3	01-02-2018	7	0	0	0	78	0	0	0	0	140	0	0	0	0	2	0	0	0	0
4	01-03-2018	11	0	20	0	139	0	0	0	0	260	0	0	0	0	5	1	0	0	2
5	01-04-2018	9	0	1	0	64	0	0	0	0	111	0	0	0	0	1	0	0	0	0
6	01-05-2018	3	0	15	0	112	0	0	0	0	173	4	0	12	0	3	0	1	1	1
7	01-06-2018	18	0	0	0	131	0	0	0	0	215	2	0	11	0	7	0	0	0	0
8	01-07-2018	-3	0	2	0	61	0	0	0	0	132	0	0	0	0	1	0	0	0	0
9	01-08-2018	10	0	0	0	82	2	0	2	0	143	0	0	0	0	2	0	0	0	0
10	01-09-2018	12	0	3	0	125	0	0	0	0	198	0	0	0	3	8	2	0	0	0
11	01-10-2018	0	4	0	0	79	0	0	0	0	165	0	0	0	0	2	0	0	0	0
12	01-11-2018	1	0	0	2	92	0	0	0	0	167	0	0	0	0	10	0	0	0	0
13	01-12-2018	11	0	9	0	183	0	0	0	0	292	0	0	6	0	16	1	0	0	0
14	01-01-2019	4	0	3	0	79	0	0	0	0	111	0	0	0	0	3	1	0	0	0
15	01-02-2019	3	0	0	0	98	1	0	1	0	76	1	0	0	0	0	0	0	0	0
16	01-03-2019	7	0	0	0	184	0	0	0	0	132	9	1	12	0	9	0	0	0	0
17	01-04-2019	1	0	0	0	48	0	0	0	0	34	3	0	0	0	0	0	0	0	0
18	01-05-2019	1	0	19	0	77	0	0	0	0	75	0	0	0	0	0	0	0	0	0
19	01-06-2019	10	0	22	0	237	0	0	0	0	110	0	1	13	0	4	0	0	0	0
20	01-07-2019	2	0	0	0	55	0	0	0	0	6	0	0	0	0	0	4	0	0	0
21	01-08-2019	1	0	0	0	60	0	0	0	0	1	0	0	0	0	1	0	0	0	0
22	01-09-2019	15	0	132	0	127	0	0	0	0	7	0	0	131	0	0	0	0	0	0
23	01-10-2019	1	0	0	0	64	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Exhibit 5

Vishal has managed to collect the data and now using ARIMA/Winter-Holt models in R SHINY has started working on the tool.

### TEACHING NOTE

#### Building forecasting model for multi varied products using R shiny at Hexagon Geo systems India

##### CASE SUMMARY

Vishal worked as an intern at Hexagon Geosystems, India wherein he was assigned a role of operations intern and specified a task to improve forecasting capabilities for a wide portfolio of products. The as-is forecasting techniques were concentrated around straight line and moving average models that give promising results when the product lies in a smooth category. However, the basic underlying problem that Vishal had to face was that 80% of the products fell under the intermittent, Erratic and Lumpy categories which are supposed to have unreliable forecasts when forecasted using the above-mentioned models. Products are categorized under 4 major clusters that significantly determine their forecasting capabilities based on Average Demand Interval and Coefficient of Variation. So, Vishal kicked off his project by categorizing these products and with a detailed exploratory analysis over the data set. For all but the smooth demand profile, forecast accuracy is not a reliable performance metric. It lacks contextual information and, in the end, leads you to miss the big picture. This induces overstock situations or, on the contrary, poor service level, both situations you want to avoid. This is why you should take some time to understand your products' various demand patterns, step back, and adjust your expectations.

It was crucial to understand products' demand patterns and optimize time and resource allocation for forecasting more than 500 products at a single click of a button. Vishal was given the task of developing a robust forecasting model that would be capable of forecasting massive number of products at minimum time allocation.

The case briefs upon various forecasting models used and deals with the process and techniques of building a forecasting model for massive number of products using interactive capabilities in R Shiny and Excel as a basic data repository.

### LEARNING OBJECTIVES

The primary objectives of this case are to discuss the use of mathematical techniques to analyze and optimize sales forecasting wide portfolio of cutting-edge technology products showcasing an intermittent nature of historical sales. Its learning objectives include:

1. Forecasting the future demand for products using ARIMA and Holt Winters' forecasting techniques. The students may be given an opportunity to explore various forecasting techniques such as moving average, exponential smoothing, etc. The instructor may ask the students to check if there is seasonality in the data.
2. Understanding the presence of seasonality in the data, and using techniques to de-seasonalize the data.
3. Understanding statistical tools like Ljung-Box test to validate the forecasting models.
4. Using Batch forecasting to generate forecasts for multiple products utilizing minimum resources like manpower and time.

### SUGGESTED CASE QUESTIONS

1. What would be the most apt multi products forecasting techniques for Hexagon and Why?
2. What data points should Vishal consider before developing the tool? What should be the outcome of Exploratory analysis?
3. Would tools be a right choice? Why?
4. If yes, how should it be developed?

### TEACHING PLAN AND TIMING

Note: Teacher should instruct students to install R Studio before attending the class

- **15 minutes:** Instructor get the class to vote about the importance of Sales forecasting and its importance in operations planning. Discussions should also be focused around the different patterns of historical demand generally witnessed in different manufacturing facilities and its impact on forecast ability.
- **15 minutes:** Assignment question 1 – the instructor should divide the class in two and ask them to come up with a suitable forecasting model for the sample product data and also categorize the sample

products based on demand intervals and variation.

- **10 minutes:** Assignment question 2 – the instructor gets the class to discuss on the best suitable model for forecasting the product samples. Teacher should provide reasonable argument to each of the answers. The two groups formed earlier discuss internally on the one evaluation method they would follow.
- **10 minutes:** Assignment question 3– the instructor should then ask the two groups formed to evaluate different statistical tools available to build a forecasting model.
- **15 minutes:** Assignment question 4–The instructor could show a YouTube clip related to the development of the forecasting model using the right statistical tool.
- **20 minutes:** Assignment question 5– The instructor then asks the students to show the different models that they have prepared along with their respective accuracies and also assists them to make a sample model, thus helping to make relevant business decision based on the forecast accuracies.
- **5 minutes:** Conclusion restating the key learning points or reading the postscript.

## SUGGESTED SOLUTIONS TO CASE QUESTIONS

### Steps involved in Forecasting

#### 1: Define the problem.

Frequently this is the most troublesome piece of gauging. Characterizing the issue cautiously requires a comprehension of the manner in which the gauges will be utilized, who requires the estimates, and how the gauging capacity fits inside the association requiring the figures. A forecaster needs to invest energy conversing with each and every individual who will be associated with gathering information, looking after

databases, and utilizing the conjectures for future arranging.

#### 2: Collect the available information.

There are consistently in any event two sorts of data required: (a) measurable information, and (b) the amassed mastery of the individuals who gather the information and utilize the conjectures. Frequently, it will be hard to acquire enough chronicled information to have the option to fit a decent factual model. All things considered; the critical gauging techniques can be utilized. Once in a while, old information will be less helpful because of basic changes in the framework being conjecture; at that point we may decide to utilize just the latest information. In any case, recollect that great factual models will deal with transformative changes in the framework; don't discard great information pointlessly.

#### 3: Perform Descriptive and exploratory analysis.

Continuously start by charting the information. Are there reliable examples? Is there a huge pattern? Is irregularity significant? Is there proof of the nearness of business cycles? Are there any anomalies in the information that should be clarified by those with master information? How solid are the connections among the factors accessible for investigation? Different instruments have been created in R to help with this examination.

#### 4: Choose data relevant fitting models.

The best model to utilize relies upon the accessibility of authentic information, the quality of connections between the figure variable and any logical factors, and the manner by which the estimates are to be utilized. It isn't unexpected to look at a few likely models. Each model is itself a fake develop that depends on a series of expectations (unequivocal and certain) and as a rule includes at least one boundary which must be evaluated utilizing the known chronicled information.

#### Step 5: Evaluate and use the model.

When a model has been chosen and its boundaries evaluated, the model is utilized to make figures. The exhibition of the model must be appropriately assessed after the information for the figure time frame have opened up. Various techniques have been created to help in evaluating the precision of conjectures. There are likewise authoritative issues in utilizing and following up on the conjectures.



### Available Forecasting Techniques

- **Simple Moving Average:** The simplest form of a moving average, known as a simple moving average (SMA), is calculated by taking the arithmetic mean of a given set of values. In other words, a set of numbers—or sales values in the case of SKUs or product level—are added together and then divided by the number of sales values in the set.
- **Exponential Moving Average:** The exponential moving average is a type of moving average that gives more weight to recent sales in an attempt to make it more responsive to new information. To calculate an EMA, you must first compute the simple moving average (SMA) over a particular time period. Next, you must calculate the multiplier for weighting the EMA (referred to as the "smoothing factor"). Then you use the smoothing factor combined with the previous EMA to arrive at the current value. The EMA thus gives a higher weighting to recent prices, while the SMA assigns equal weighting to all values.
- **Holt-Winters Forecasting:** Holt-Winters forecasting is a way to model and predict the behavior of a sequence of values over time—a time series. Holt-Winters is one of the most popular forecasting techniques for time series. It's decades old, but it's still ubiquitous in many applications, including monitoring, where it's used for purposes such as anomaly detection and capacity planning. Holt-Winters is a model of time series behavior. Forecasting always requires a model, and Holt-Winters is a way to model three aspects of the time series: a typical value (average), a slope (trend) over time, and a cyclical repeating pattern (seasonality). Holt-Winters uses exponential smoothing to encode lots of values from the past and use them to predict "typical" values for the present and future.

The three aspects of the time series behavior—value, trend, and seasonality—are expressed as three types of exponential smoothing, so Holt-Winters is called triple exponential smoothing. The model predicts a current or future value by

computing the combined effects of these three influences. The model requires several parameters: one for each smoothing ( $\alpha$ ,  $\beta$ ,  $\gamma$ ), the length of a season, and the number of periods in a season.

Seasonality can be confusing. A season is a fixed length of time that contains the full repetition. You might think your data repeats daily (there's a peak at 2pm every day), but if the weekend has different behavior (there's no peak at 2pm on Sunday) then your season is really a week, not a day. Within the season, there are periods, which is the granularity of prediction. If you want to model a value for every hour of every day within a week, your season is 168 hours long and your period is 1 hour.

- **ARIMA Forecasting:** ARIMA, short for 'Auto Regressive Integrated Moving Average' is actually a class of models that 'explains' a given time series based on its own past values, that is, its own lags and the lagged forecast errors, so that equation can be used to forecast future values. Any 'non-seasonal' time series that exhibits patterns and is not a random white noise can be modeled with ARIMA models. An ARIMA model is characterized by 3 terms: p, d, q where,

p is the order of the AR term

q is the order of the MA term

d is the number of differencing required to make the time series stationary

If a time series, has seasonal patterns, then you need to add seasonal terms and it becomes SARIMA, short for 'Seasonal ARIMA'. The first step to build an ARIMA model is to [make the time series stationary](#). Because, term 'Auto Regressive' in ARIMA means it is a [linear regression model](#) that uses its own lags as predictors. Linear regression models, as you know, work best when the predictors are not correlated and are independent of each other. So how to make a series stationary?

The most common approach is to difference it. That is, subtract the previous value from the current value. Sometimes, depending on the complexity of the series, more than one differencing may be needed. The value of d, therefore, is the minimum number of differencing needed to make the series stationary. And if the time series is already stationary, then  $d = 0$ .

'p' is the order of the 'Auto Regressive' (AR) term. It refers to the number of lags of Y to be used as predictors. And 'q' is the order of the 'Moving Average' (MA) term. It refers to the number of lagged forecast errors that should go into the ARIMA Model.

**ARIMA model in words:** Predicted  $Y_t$  = Constant + Linear combination Lags of Y (upto p lags) + Linear Combination of Lagged forecast errors (upto q lags)

library(forecast)

retail <- read.csv ("FILE NAME", header=TRUE/FALSE)

retail <- ts (retail [,1], f=12, s=1982+3/12)

ns <- ncol (retail)

h <- 24

fcast <- matrix (NA, nrow=h, ncol=ns)

for (i in 1:ns)

fcast [,i] <- forecast(retail[,i], h=h)\$mean

write(t(fcast),file="retailfcasts.csv", sep="," , ncol=ncol(fcast))

### Data Gathering and Exploratory Analysis

The initial fundamental steps for building a robust forecasting model are quality data collection and its exploratory analysis.

In this case the data requirement should be for at least past 2 years of sales data aggregated on a monthly basis. Data must be collected from trustworthy sources and should not include any biases. The next step is to perform exploratory analysis with includes ABC/XYZ analysis, trend decompositions and correlation analysis. The following sections will describe about the mentioned analysis in detail.

### ABC & XYZ Analysis

ABC/XYZ examination is a technique for gathering arranging objects (trademark esteem mixes, SKUs) in view of their worth (income or deals volume) and elements of utilization or deals. During the examination, the arranging objects are relegated one of the classes of ABC and XYZ all the while.

Investigation dependent on esteem (ABC examination) should be possible by applying

### Multiple Product Forecasting Scenario

Loops and data combining methods can be used in are to run forecasting models over a wide range of products. The same can be leveraged to compare the forecasting accuracies dynamically and assign the best fit model for a specific product.

Following the example of looping code that students can use to run on multiple products:

Reads the input file

Reads the input file

Assigns the number of products to variable "ns"

Sets a forecasting horizon of 24 months in future

Runs the forecasts for every product then prints the point forecasts as results in a csv file format and saves it in the working directory

Pareto's standard on the estimation of individual things in a particular time skyline. The examination positions arranging objects as indicated by their worth commitments as far as deals or commitment edge. Investigation dependent on the elements of utilization (XYZ examination) implies the figuring of a change coefficient of utilization or deals in a particular time skyline. For instance, an arranging object that has the qualities BY might mean an arranging object with medium volume, and pattern or occasional interest.

Using this tool Vishal observed that the majority of high value products (A class) were placed in a less reliable (Erratic) forecasting domain. For medium value products (B class) half of them were placed in Y segment (Intermittent) and 43% of them were in the less reliable segment. It is interesting to notice that 96% of the products from the low value (C class) segment have an intermittent type of demand. A maximum of 17% of the A class products have a smooth demand pattern and 7% in the case of B class products.

	A	B	C
X	<b>AX Class</b> <ul style="list-style-type: none"><li>• High consumption value.</li><li>• Even demand.</li><li>• Reliable forecasts.</li></ul>	<b>BX Class</b> <ul style="list-style-type: none"><li>• Medium consumption value.</li><li>• Even demand.</li><li>• Reliable forecasts.</li></ul>	<b>CX Class</b> <ul style="list-style-type: none"><li>• Low consumption value.</li><li>• Even demand.</li><li>• Reliable forecasts.</li></ul>
Y	<b>AY Class</b> <ul style="list-style-type: none"><li>• High consumption value.</li><li>• Predictably variable demand.</li><li>• Less reliable forecasts.</li></ul>	<b>BY Class</b> <ul style="list-style-type: none"><li>• Medium consumption value.</li><li>• Predictably variable demand.</li><li>• Less reliable forecasts.</li></ul>	<b>CY Class</b> <ul style="list-style-type: none"><li>• Low consumption value.</li><li>• Predictably variable demand.</li><li>• Less reliable forecasts.</li></ul>
Z	<b>AZ Class</b> <ul style="list-style-type: none"><li>• High consumption value.</li><li>• Sporadic, variable demand.</li><li>• Forecasting unreliable or impossible.</li></ul>	<b>BZ Class</b> <ul style="list-style-type: none"><li>• Medium consumption value.</li><li>• Sporadic, variable demand.</li><li>• Forecasting unreliable or impossible.</li></ul>	<b>CZ Class</b> <ul style="list-style-type: none"><li>• Low consumption value.</li><li>• Sporadic, variable demand.</li><li>• Forecasting unreliable or impossible.</li></ul>

Exhibit 1

Time Series Patterns

In portraying these time arrangements, we have utilized words, for example, "pattern" and "occasional" which should be characterized all the more cautiously.

Pattern

An example exists when there is a drawn-out addition or reducing in the data. It shouldn't be straight. A portion of the time we will simply an example as "changing course", when it might go from a growing example to a reducing design.

Occasional

An intermittent model happens when a period course of action is affected by means of incidental components, for instance, the season or the day of the week. Anomaly is reliably of a fixed and known repeat.

Cyclic

A cycle happens when the data show rises and falls that are not of a fixed repeat. These differences are generally a direct result of

monetary conditions, and are routinely related to the "business cycle". The length of these fluctuations is generally speaking in any occasion 2 years.

Assumptions in Time Series forecasting

Autocorrelation is a sort of sequential reliance. In particular, autocorrelation is the point at which a period arrangement is straight identified with a slacked variant of itself. On the other hand, connection is essentially when two free factors are straight related.

For repetitive sound, we anticipate that every autocorrelation should be near zero. It is entirely expected to plot these limits on a chart of the ACF (the blue ran lines above). In the event that at least one enormous spike is outside these limits, or in the event that considerably over 5% of spikes are outside these limits, at that point the arrangement is most likely not repetitive sound.

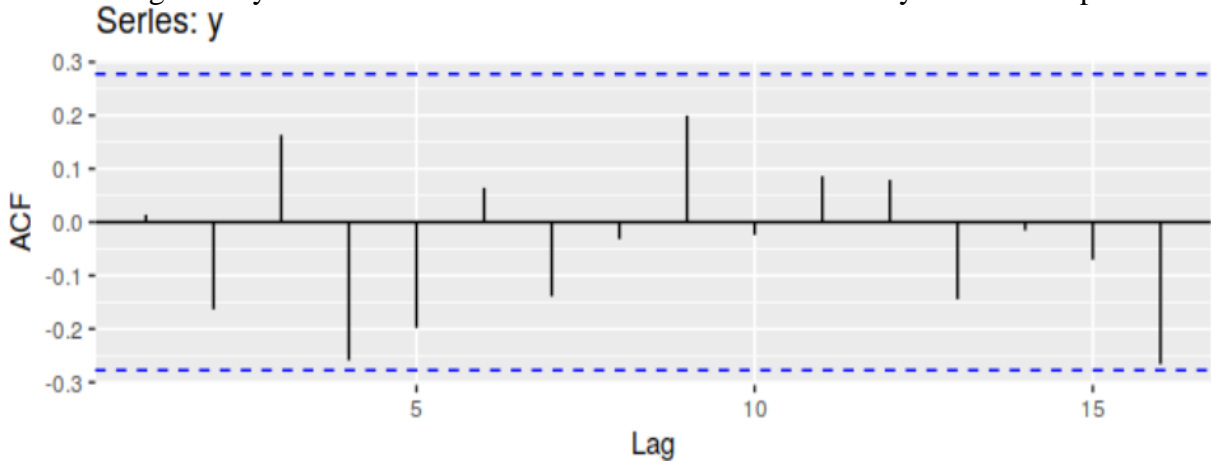


Exhibit 2

In this model,  $T=50$  thus the limits are at  $\pm 2/\sqrt{50}=\pm 0.28$ . The entirety of the autocorrelation coefficients exists in these cutoff points, affirming that the information are background noise. Autocorrelation is a sort of sequential reliance. In particular, autocorrelation is the point at which a period arrangement is straightly identified with a slacked variant of itself. On the other hand, connection is essentially when two free factors are straightly related. In this model,  $T=50$  thus the limits are at  $\pm 2/\sqrt{50}=\pm 0.28$ . The entirety of the

autocorrelation coefficients exists in these cut-off points, affirming that the information are background noise. Utilize the incomplete autocorrelation and autocorrelation works together to recognize ARIMA models. Search for the accompanying examples on the incomplete autocorrelation work. Analyze the spikes at each slack to decide if they are centrality. A critical spike will reach out past as far as possible, which demonstrates that the connection for that slack doesn't rise to zero.




Pattern	Indicates	Example
Large spike at lag 1 that decreases after a few lags.	A moving average term in the data. Use the autocorrelation function to determine the order of the moving average term.	
Large spike at lag 1 followed by a damped wave that alternates between positive and negative correlations.	A higher order moving average term in the data. Use the autocorrelation function to determine the order of the moving average term.	
Significant correlations at the first or second lag, followed by correlations that are not significant.	An autoregressive term in the data. The number of significant correlations indicate the order of the autoregressive term.	

Exhibit 3

Testing the Model

Residuals are important in checking whether a model has enough gotten the information in the data. An average envisioning system will yield residuals with the going with properties:

1. The residuals are uncorrelated. On the off chance that there are connections between residuals, at that point there is data left in the residuals which ought to be utilized in processing figures.
2. The residuals have zero mean. In the event that the residuals have a mean other than zero, at that point the gauges are one-sided.

Any assessing technique that doesn't satisfy these properties can be improved. Regardless, that doesn't suggest that gauging methods that satisfy these properties can't be improved. It is possible to have a couple of differing gauging systems for comparative enlightening file, all of which satisfy these properties. Checking these properties is huge in order to see whether a procedure is using the total of the available information, yet it's definitely not a not too bad technique to pick a deciding methodology. If both of these properties aren't satisfied, by then the gauging procedure can be changed to give

better guesses. Changing for tendency is basic: if the residuals have mean mm, by then basically add mm to all measures and the inclination issue is comprehended. Despite these crucial properties, it is useful (yet repetitive) for the residuals to in like manner have the going with two properties.

3. The residuals have steady change.
4. The residuals are regularly dispersed.

Time Series Decomposition

STL is a flexible and hearty technique for breaking down time arrangement. STL is an abbreviation for "Occasional and Trend decay utilizing Loess", while Loess is a strategy for evaluating nonlinear connections. The STL technique was created by Cleveland, Cleveland, McRae, and Terpenning in 1990. Then again, STL has a few inconveniences. Specifically, it doesn't deal with exchanging day or schedule variety naturally, and it just gives offices to added substance deteriorations. It is conceivable to get a multiplicative disintegration by first taking logs of the information, at that point back-changing the parts. Deteriorations among added substance and multiplicative can be gotten utilizing a Box-Cox change of the information with  $0<\lambda<10<\lambda<1$ . An estimation of  $\lambda=0$  relates to the multiplicative



decay while  $\lambda=1$  is comparable to an added substance disintegration.

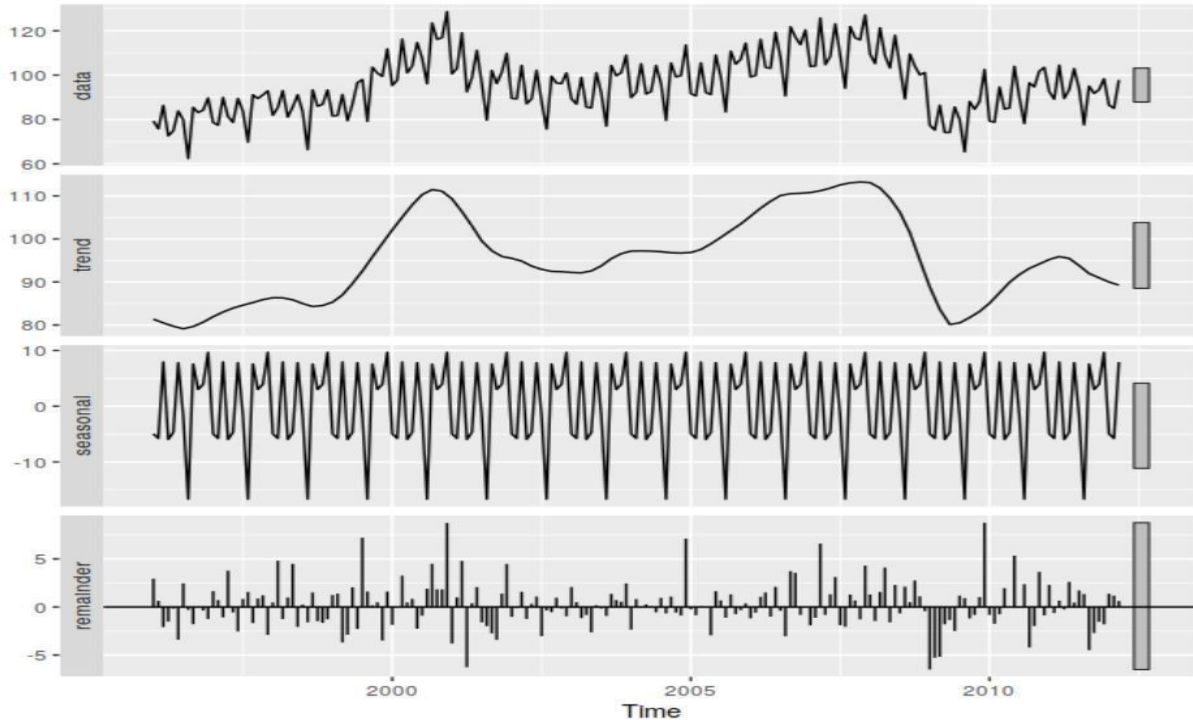


Exhibit 4

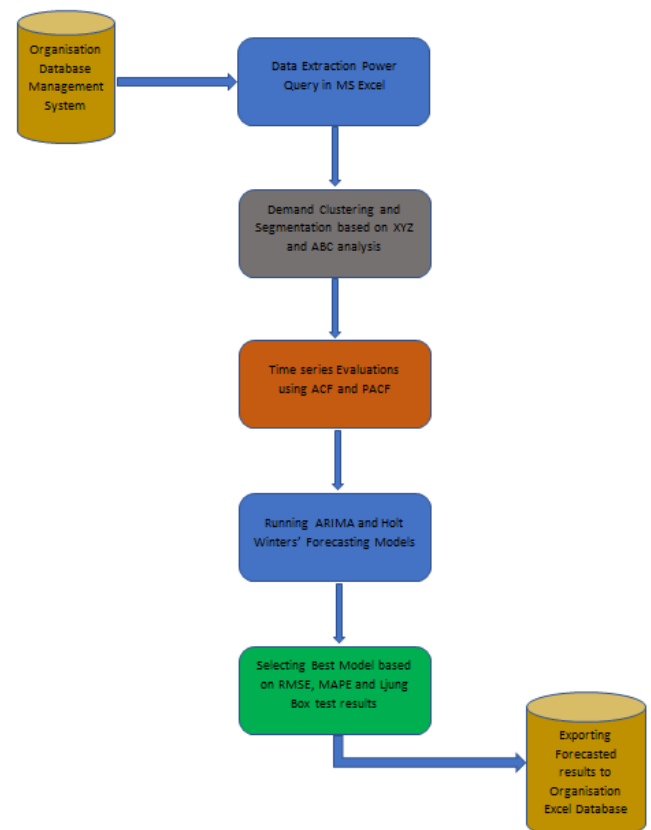
### Why developing a tool is the right choice? How should it be deployed?

A robust forecasting tool developed in R Shiny will help the user to engage directly with the interface without worrying about the background scripts. It will give the user a certain amount of flexibility to compare different models and understand basic trend, seasonality and structure of the historical data pattern.

The tool should be designed after rigorous brainstorming with the demand planning executives working in the organization. All the requirements must be taken into consideration and a detailed view upon future modifications in the product portfolio must be understood prior developing the tool.

During the deployment phase a detailed user manual must be circulated among the targeted users and a focused time frame must be assigned for a detailed user training on the tool.

### Flow Chart for Forecasting Tool Development



Tool Interface and Fundamentals

Inside most associations, the words "request arranging" cause a response—and normally not a gentle one. It is portrayed by enthusiastic limits like indignation, depression, bafflement, or even sadness. Only here and there do we discover a group energized or hopeful about their odds to improve request arranging forms. Following two many years of procedure and innovation refinement, greatness popular administration

despite everything escapes gracefully chain groups. Truth be told, it is the flexibly chain arranging application with the best hole among execution and fulfillment. Simultaneously, it's the application with the best arranged future spending. For most groups, request arranging is a problem, a genuine romance detest relationship. They need to improve the interest arranging process, however stay suspicious that they can ever do as such.

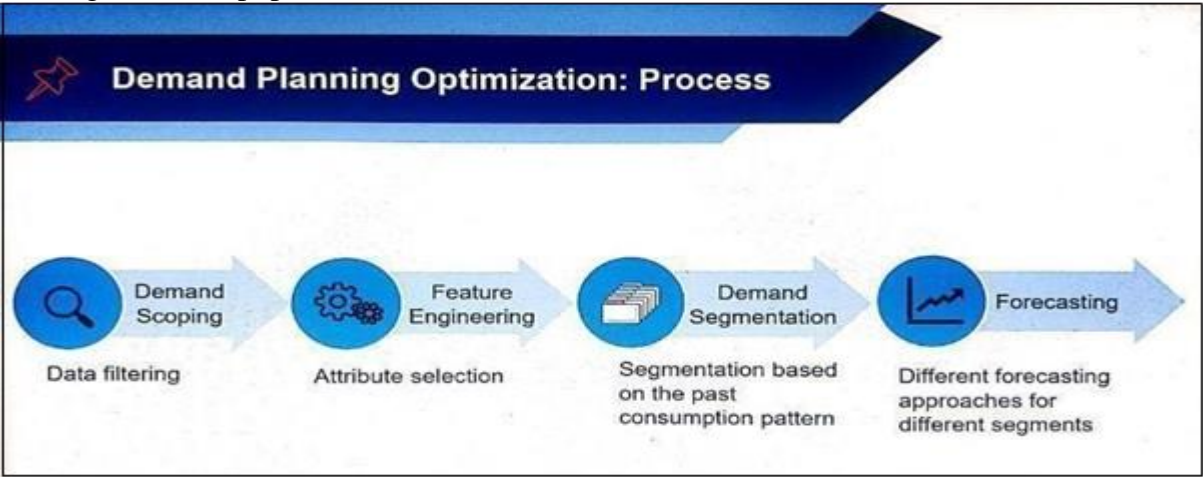


Exhibit 6

As talked about in the previous segments, the means before estimating have just been finished. In the accompanying areas we will talk about the coordinated R gleaming application that bolsters any interest organizer at the stockroom to proficiently estimate the future deals of any item from the information base. This apparatus utilizes exceed expectations to produce beginning essential conjectures like line of best fit, exponential smoothening techniques. The exceed expectations record likewise empowers the client

to produce the R sparkling information document that is taken care of to glossy application where increasingly complex determining techniques like Holt Winters and ARIMA are utilized. Inside most associations, the words "request arranging" cause a response—and normally not a gentle one. It is portrayed by enthusiastic limits like indignation, depression, bafflement, or even sadness. Only here and there do we discover a group energized or hopeful about their odds to improve request arranging forms.

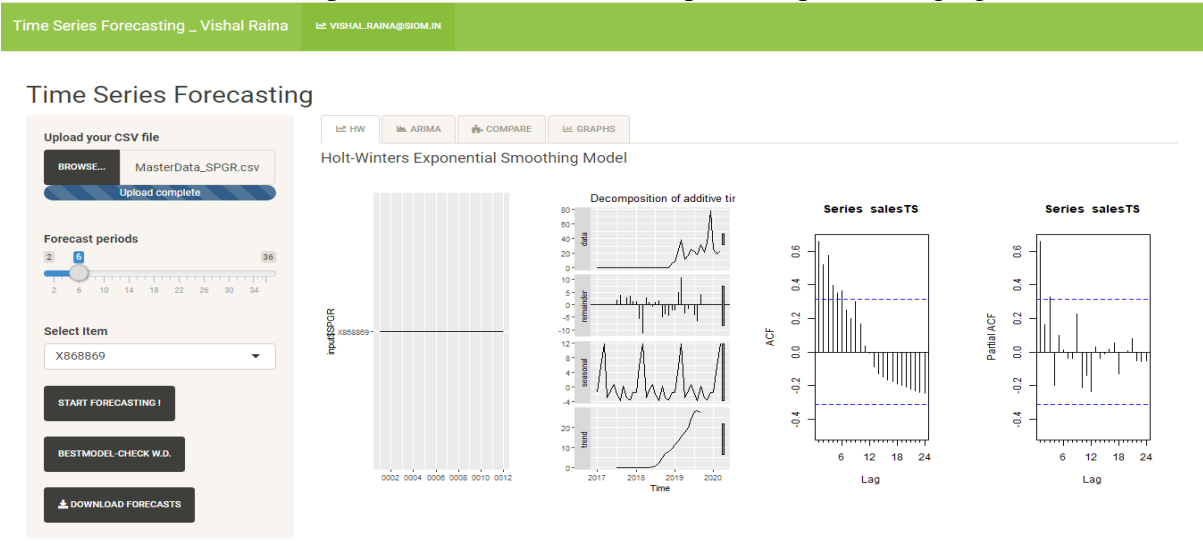


Exhibit 7

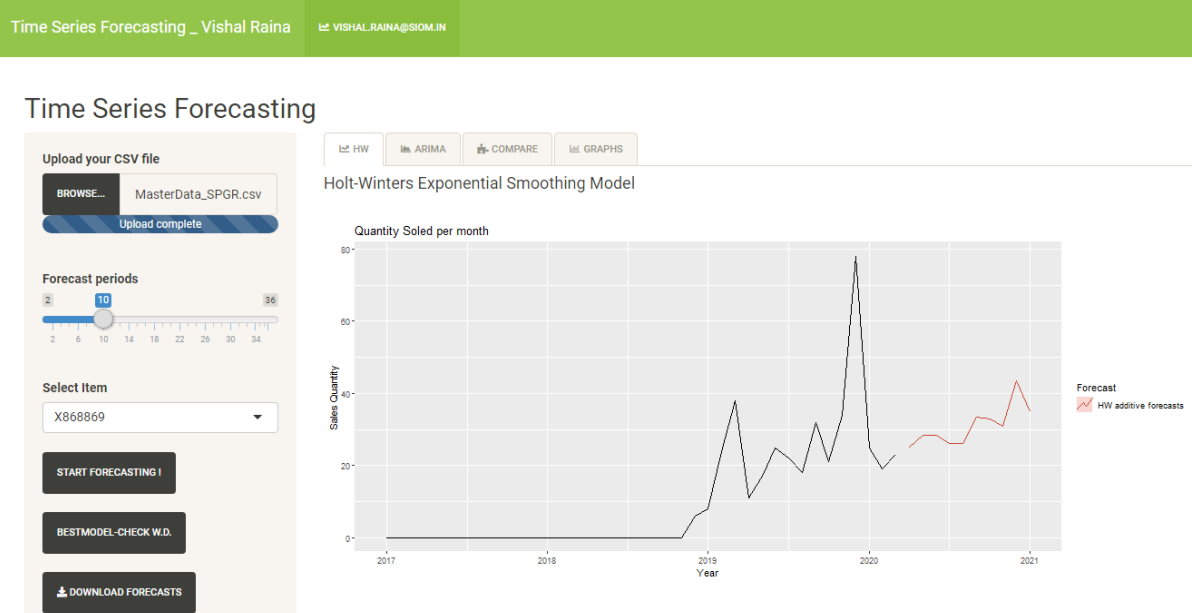
This picture describes the initial interface of the forecasting tool. This allows the user to upload the respective csv file where the historical sales figures have been saved. It also gives the flexibility to select the forecasting horizon within a range of 2 to 36 months using the slide bar. Below the slide bar there is a selection box, wherein the user can select the product code of the desired item to be forecasted. After providing the above-mentioned inputs the user has to select the “START FORECASTING!” button to populate the results in the side tab.

Initially the side tab represents the seasonal decomposition, ACF and PCF plots for the selected product that helps the user to understand the Trend, Seasonality, Cyclicity and also judge whether the time series is a White Noise or a forecastable pattern.

The Shiny app has multiple functionalities and is very user friendly. It allows the user to upload the recent R input file generated from the excel interface. The user can also set the forecasting periods by sliding the button on the slider placed below the Browse section. Once the file has been

uploaded the user can view the data set and time series decomposition plots in the main panel. This helps the user to verify the data before forecasting also helps in analyzing the seasonality, trend and cyclic behavior in data. As discussed in the previous sections, the user can also check the respective plots to verify the forecast ability of the time series. Once the user presses the Start Forecasting button the app initiates the ML algorithm and prints out results under the HW and ARIMA side tabs. These are easily accessible and understandable for the user. After analyzing the results, user can download the forecasts under the 95% confidence interval by clicking on the download forecast button.

The user can observe the forecasting trend in red colored line plotted in the graph shown. Under the performance metrics panel, user can check the accuracy summary for respective forecasting models as well as the hypothesis testing (Box-Ljung) results to validate the model.



**Exhibit 8**

This interface represents the forecasted result using Winter Holtz forecasting method. The next

tab represents the forecasted result using ARIMA method.

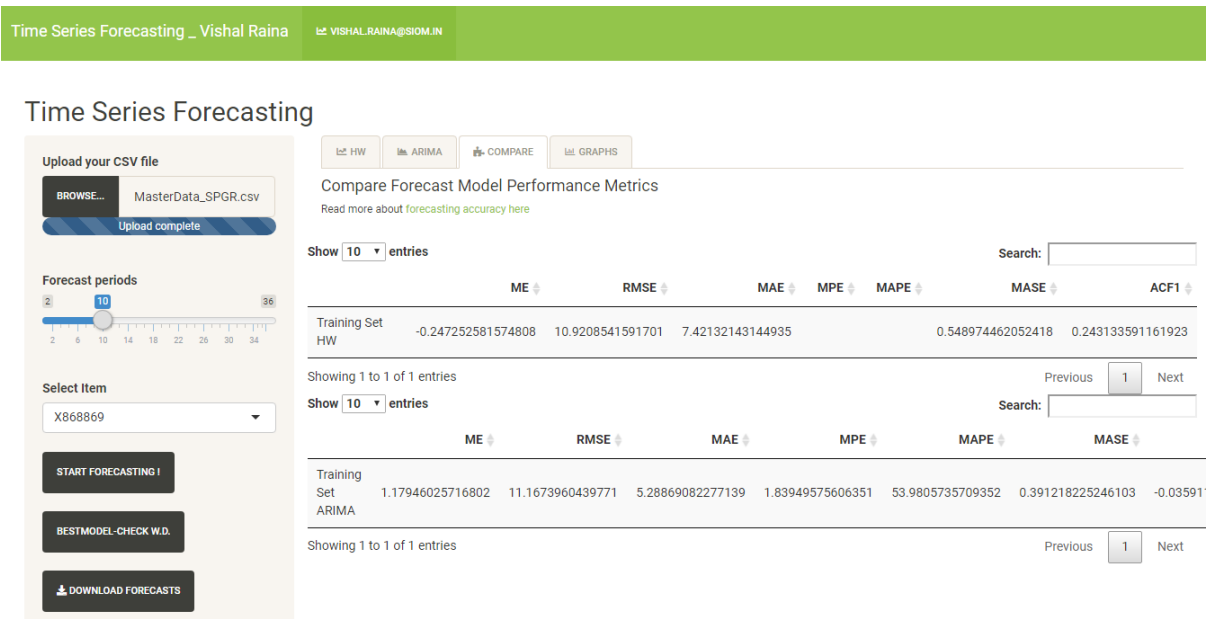
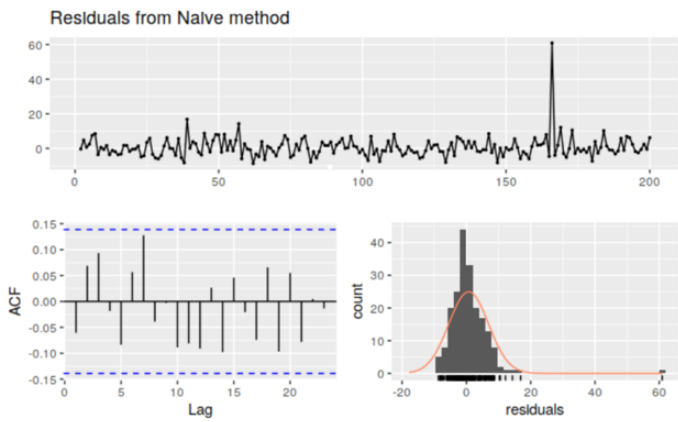


Exhibit 9

Under the Compare tab the user can check the feasibility of the forecasting models by comparing their RMSE and MAPE values. The model with

lower error values must be considered for further analysis.



Under the performance metrics panel, user can check the accuracy summary for respective forecasting models as well can the hypothesis testing (Ljung-Box) results to validate the model.

**H0:** The data are independently distributed (i.e. the correlations in the population from which the sample is taken are 0, so that any observed correlations in the data result from randomness of the sampling process).

**Ha:** The data are not independently distributed; they exhibit serial correlation.

Exhibit 10

The Performance Metrics section under each model tab represents the above mentions parameters that can be leveraged by the user to

check if the model is statistically significant or incorporated some discrepancies.



Time Series Forecasting

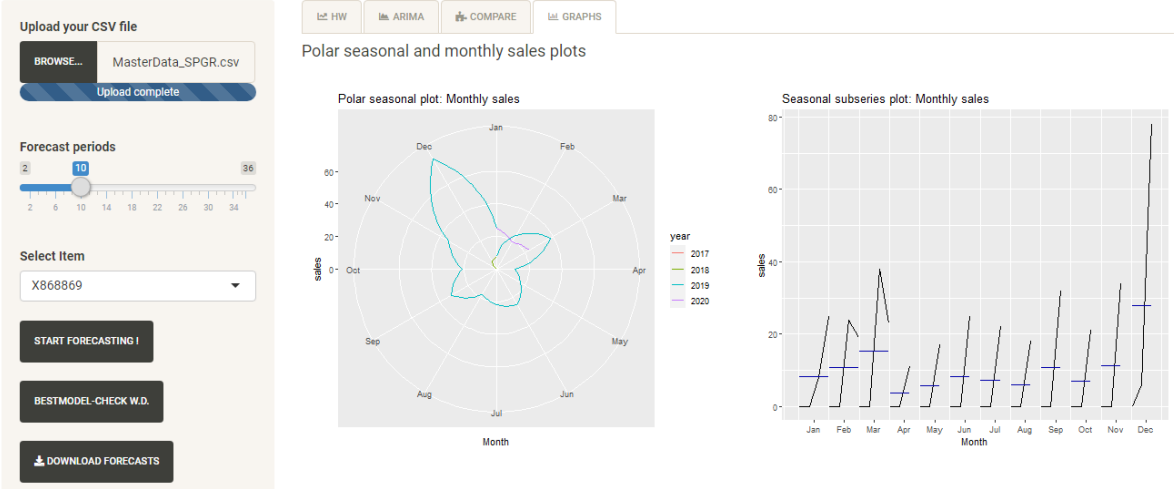


Exhibit 10

Under the “Graphs” tab the user can infer about most active sales periods as well as the seasonal trends categorized at monthly levels.

Tool Forecasting vs Moving average Forecasting

In this section we will compare the performance of this forecasting tool with a moving average model for a sample product data that was primarily used in the organization. This will help to build confidence in the time series analysis models using exponential smoothening and integrated liner regression methods.

Month	Historical Data for Sample Product	3 Week Moving Average	RMSE	MAPE
Jan-18	0		3.525043	68%
Feb-18	0			
Mar-18	0			
Apr-18	0	0	0	
May-18	0	0	0	
Jun-18	0	0	0	
Jul-18	0	0	0	
Aug-18	0	0	0	
Sep-18	0	0	0	
Oct-18	0	0	0	
Nov-18	0	0	0	
Dec-18	7	0	-7	1
Jan-19	4	2.333333333	-1.666666667	0.416666667
Feb-19	3	3.666666667	0.666666667	0.222222222
Mar-19	15	4.666666667	-10.33333333	0.688888889
Apr-19	2	7.333333333	5.333333333	2.666666667
May-19	4	6.666666667	2.666666667	0.666666667
Jun-19	6	7	1	0.166666667
Jul-19	5	4	-1	0.2
Aug-19	2	5	3	1.5
Sep-19	7	4.333333333	-2.666666667	0.380952381
Oct-19	6	4.666666667	-1.333333333	0.222222222
Nov-19	9	5	-4	0.444444444
Dec-19	13	7.333333333	-5.666666667	0.435897436
Jan-20	4	9.333333333	5.333333333	1.333333333
Feb-20	9	8.666666667	-0.333333333	0.037037037
Mar-20	6	8.666666667	2.666666667	0.444444444

Exhibit 11

Now, we will check the accuracy values for the same sample product using the R Shiny tool. We will observe that the tool provides lower error

values and parallelly helps the user to select the best model among the listed ones.

In the current case this tool has helped to reduce RMSE forecasting error by 15% using the Holt Winters method and by 8.5 % using the ARIMA

method. Hence the user must select the Holt Winters for this specific product.

Time Series Forecasting

Upload your CSV file

BROWSE...

MasterData\_SPGR.csv

Upload complete

Forecast periods

3

36

Select Item

X868851

START FORECASTING !

BESTMODEL-CHECK W.D.

DOWNLOAD FORECASTS

HWARIMACOMPAREGRAPHS

Compare Forecast Model Performance Metrics

Read more about forecasting accuracy here

Show 10 entries

ME

RMSE

MAE

MPE

MAPE

Training Set

HW

0.0408242082413508

2.95728628234568

1.93613531550766

Showing 1 to 1 of 1 entries

Show 10 entries

ME

RMSE

MAE

MPE

Training Set

ARIMA

0.495275545581498

3.22521164780312

2.10896402621272

-15.0977759657991

Showing 1 to 1 of 1 entries

Sample Data for Practice

Date	Product 1	Product 2	Product 3	Product 4
01-01-2018	42	104	279	0
01-02-2018	7	78	140	0
01-03-2018	11	139	260	0
01-04-2018	9	64	111	0
01-05-2018	3	112	173	12
01-06-2018	18	131	215	11
01-07-2018	-3	61	132	0
01-08-2018	10	82	143	0
01-09-2018	12	125	198	0
01-10-2018	0	79	165	0
01-11-2018	1	92	167	0
01-12-2018	11	183	292	6
01-01-2019	4	79	111	0
01-02-2019	3	98	76	0
01-03-2019	7	184	132	12
01-04-2019	1	48	34	0
01-05-2019	1	77	75	0
01-06-2019	10	237	110	13
01-07-2019	2	55	6	0
01-08-2019	1	60	1	0
01-09-2019	15	127	7	131

01-10-2019	1	64	0	0
01-11-2019	0	137	10	2
01-12-2019	26	261	24	2
01-01-2020	2	80	2	0
01-02-2020	1	103	8	0
01-03-2020	23	146	5	20

## References

- [1] Amin-Naseri, M. R., & Tabar, B. R. (2008). Neural network approach to lumpy demand forecasting for spare parts in process industries. 2008 International Conference on Computer and Communication Engineering, 1378–1382. <http://doi.org/10.1109/ICCCE.2008.4580831>
- [2] Chen, F.-L., & Chen, Y.-C. (2009). An Investigation of Forecasting Critical Spare Parts Requirement. In 2009 WRI World Congress on Computer Science and Information Engineering (Vol. 4, pp. 225–230). IEEE. <http://doi.org/10.1109/CSIE.2009.774>
- [3] Croston, J. D. (1972). Forecasting and Stock Control for Intermittent Demands. *Operational Research Quarterly* (1970-1977), 23(3), 289–303. <http://doi.org/10.2307/3007885>
- [4] Eaves, A. H. C., & Kingsman, B. G. (2004). Forecasting for the ordering and stock-holding of spare parts. *Journal of the Operational Research Society*, 55(4), 431–437. <http://doi.org/10.1057/palgrave.jors.2601697>
- [5] Ghobbar, A. A., & Friend, C. H. (2003). Evaluation of forecasting methods for intermittent parts demand in the field of aviation: A predictive model. *Computers and Operations Research*, 30(14), 2097–2114. [http://doi.org/10.1016/S0305-0548\(02\)00125-9](http://doi.org/10.1016/S0305-0548(02)00125-9)
- [6] Hua, Z., & Zhang, B. (2006). A hybrid support vector machines and logistic regression approach for forecasting intermittent demand of spare parts. *Applied Mathematics and Computation*, 181(2), 1035–1048. <http://doi.org/10.1016/j.amc.2006.01.064>
- [7] Muñoz, D. F., & Muñoz, D. F. (2011). Service Parts Management: Demand Forecasting and Inventory Control. In N. Altay & A. L. Litteral (Eds.), (pp. 105–123). London: Springer London. [http://doi.org/10.1007/978-0-85729-039-7\\_5](http://doi.org/10.1007/978-0-85729-039-7_5)
- [8] Romeijnnders, W., Teunter, R., & van Jaarsveld, W. (2012). A two-step method for forecasting spare parts demand using information on component repairs. *European Journal of Operational Research*, 220(2), 386–393. <http://doi.org/10.1016/j.ejor.2012.01.019>
- [9] Johnston, F.R. and Boylan, J.E. (1996), “Forecasting for items of intermittent demand”, *Journal of the Operational Research Society*, Vol. 47 No. 1, pp. 113–121.
- [10] Johnston, F.R., Boylan, J.E. and Shale, E.A. (2003), “An examination of the size of orders from customers, their characterization and the implications for inventory control of slow moving items”, *Journal of the Operational Research Society*, Vol. 54 No. 8, pp. 833–837.
- [11] Law, R. and Au, N. (1999), “A neural network model to forecast Japanese demand for travel to Hong Kong”, *Tourism Management*, Vol. 20 No. 1, pp. 89–97.
- [12] Wallström, P. and Segerstedt, A. (2010), “Evaluation of forecasting error measurements and techniques for intermittent demand”, *International Journal of Production Economics*, Vol. 128 No. 2, pp. 625–636.
- [13] Al-Saba, T. and El-Amin, I.M. (1999), “Artificial neural networks as applied to long-term demand forecasting”, *Journal of Artificial Intelligence in Engineering*, Vol. 13 No. 2, pp. 189–197.

- [14] Kalchschmidt, M., Zotteri, G. and Verganti, R. (2003), "Inventory management in a multi-echelon spare parts supply chain", *International Journal of Production Economics*, Vol. 81 No. 1, pp. 397-413.
- [15] Teunter, R. and Duncan, L. (2009), "Forecasting intermittent demand: a comparative study", *Journal of the Operational Research Society*, Vol. 60 No. 3, pp. 321-329.
- [16] Willemain, T.R., Smart, C.N., Schwarz, H.F., 2004. A new approach to forecasting intermittent demand for service parts inventories. *International Journal of Forecasting* 20, 375–387.
- [17] Campbell, H.S., 1963. The relationship of resource demands to air base operations. RAND, Santa Monica.