Ashwini V

Diagram

Description automatically generated

Capstone Project

Sales Forecasting

Contents

[Introduction 3](#_Toc82360935)

[Problem Statement 3](#_Toc82360936)

[Need of the study/project 3](#_Toc82360937)

[Understanding business/social opportunity 3](#_Toc82360938)

[**EDA and Business Implication** 3](#_Toc82360939)

[Understanding how data was collected in terms of time, frequency and methodology 3](#_Toc82360940)

[Visual inspection of data (rows, columns, descriptive details) and attributes 4](#_Toc82360941)

[Description 4](#_Toc82360942)

[Univariate analysis (distribution and spread for every continuous attribute, distribution of data in categories for categorical ones) 4](#_Toc82360943)

[City distribution 6](#_Toc82360944)

[Product Distribution 7](#_Toc82360945)

[Box plot for Sales and product 8](#_Toc82360946)

[Percentage of Sales and Product 10](#_Toc82360947)

[Percentage of Sales and City 11](#_Toc82360948)

[City and product and sales percentage 12](#_Toc82360949)

[Date and city and sales and product 14](#_Toc82360950)

[Cooler vs Cities vs Date 14](#_Toc82360951)

[Dry Iron vs Cities vs Date 15](#_Toc82360952)

[Food Processor vs Cities vs Date 17](#_Toc82360953)

[Oven toaster grill vs Cities vs Date 23](#_Toc82360954)

[SECF vs Cities vs Date 24](#_Toc82360955)

[Steam Iron vs Cities vs Date 26](#_Toc82360956)

[Water Heaters vs Cities Vs Date 27](#_Toc82360957)

[Missing Data and duplicates 29](#_Toc82360958)

[Outlier treatment 29](#_Toc82360959)

[Business Insights 29](#_Toc82360960)

[**Model building** 30](#_Toc82360961)

[Forecasting from the modelling results 31](#_Toc82360962)

[Mixers – Mumbai 31](#_Toc82360963)

[Mixers – Kolkata Mixers – Bangalore 32](#_Toc82360964)

[Mixers – Hyderabad 32](#_Toc82360965)

[Coolers – Chennai 33](#_Toc82360966)

[Coolers- Mumbai 33](#_Toc82360967)

[Coolers – Cochin 33](#_Toc82360968)

[Coolers – Bangalore 34](#_Toc82360969)

[Seasonal Forecast with Constant Forecast. 34](#_Toc82360970)

[Dry Iron – Mumbai Dry Iron – Kolkata 34](#_Toc82360971)

[Dry Iron – Hyderabad Dry Iron – Patna 35](#_Toc82360972)

[Recommendations and Insights 40](#_Toc82360973)

[Explore Growth Opportunities 40](#_Toc82360974)

[Sustain and keep the growth upward 41](#_Toc82360975)

[Need Attention 41](#_Toc82360976)

# Introduction

## Problem Statement

Sales play a pivotal role in a company to gain profit. Each product in an appliance company might sell based on different aspects on that particular product. We can apply different marketing strategies to sell the products. So this projects involves applying Time series forecasting techniques to predict the sales for the upcoming years to curb and increase the production of products and gain profit. And also provide business insights to the company such as production of products, sales in the near future and stocks review for upcoming seasons.

## Need of the study/project

Sales is the most important aspect of a company. All companies ought to find effective business strategies to persuade customers to buy their products. However, sales might go down and up based on different aspects such as weather condition, customer’s usage and product performance. We need to identify product sales behaviour over a period of time and forecast the sales in the upcoming years to help the production of products, drop the product from the list or increase the production according to the season. Time series forecasting algorithm have a pretty good performance on purpose to predict the sales of the different products in different cities for a period of time

## Understanding business/social opportunity

Understanding the Business/Social Opportunity Forecasting the sales with accuracy helps the appliance company to have better clarity on existing and prospective customers. They can accordingly increase or diminish their production of different products. If the company produces excess product, but the sales of that particular product is down, then that particular product remain in the store indefinitely. As a repercussion, company will incur production cost loss and product sale loss. So forecasting sales will help the company in maintaining good repute and increased revenues.

# **EDA and Business Implication**

## Understanding how data was collected in terms of time, frequency and methodology

The dataset indicates different home appliance products from different states of India. There are 10 products available with demographic variables – Sales of products from different cities and sales of each product from April 2014 till March 2018 on the 1st of every month for each city.

## Visual inspection of data (rows, columns, descriptive details) and attributes

The dataset contains 4800 observations spread across 4 variables. Product, City are categorical variable. Date and Sales are continuous variables. Each product has 480 records varying in different months of the year.

|  |  |
| --- | --- |
| Variables |  |
| Products | There are 10 products |
| Date | Date range between 2014 to 2018 |
| City | There are 10 cities for each product |
| Sales | Sales of each product in each city |

### Description

|  |  |
| --- | --- |
| **Sales** |  |
| **count** | 4.80E+03 |
| **mean** | 7.42E+05 |
| **std** | 1.31E+06 |
| **min** | 0.00E+00 |
| **25%** | 1.17E+05 |
| **50%** | 3.08E+05 |
| **75%** | 7.80E+05 |
| **max** | 1.59E+07 |

There are total of 4800 rows with 50% of 308000 and maximum of 15900000. Clearly there is outlier. But outlier cannot be treated in continuous forecasting of sales. Accurate results cannot be achieved.

## Univariate analysis (distribution and spread for every continuous attribute, distribution of data in categories for categorical ones)

Note : date has been changed to datetimeformat and Month- Date- Year format for further analysis

Bar Diagram of Sales

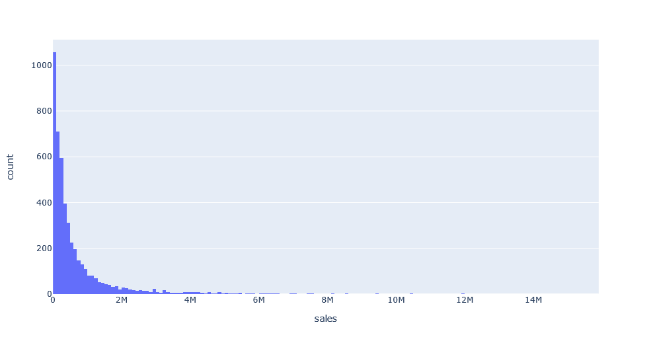


Sales variable is continuous and have outliers in the Box plot fig shown above.

Min is 0 and max outlier is 15.88M.

Outlier is not treated for sales variables because forecasting the sales data will get compromised for products during the analysis and accurate results cannot be achieved.

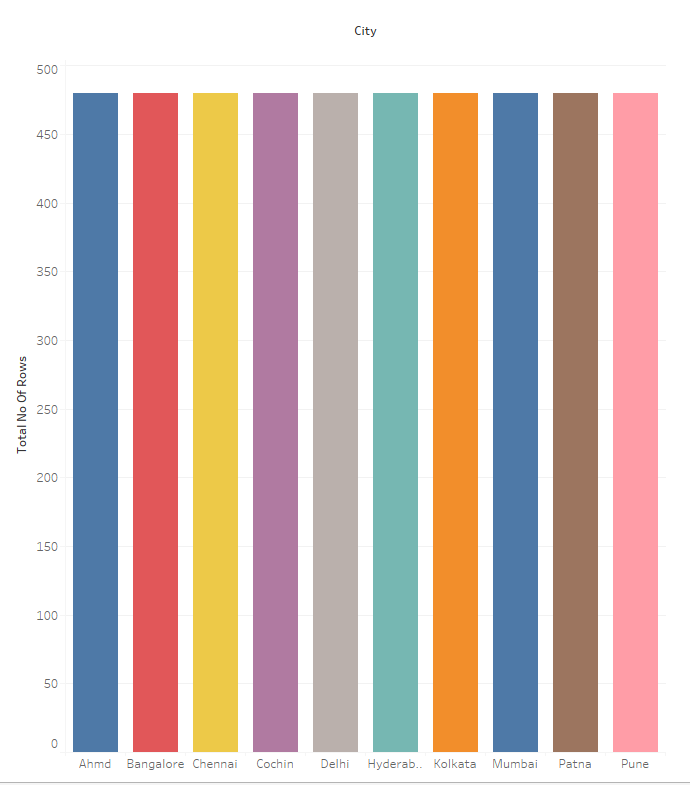
Histogram of Sales



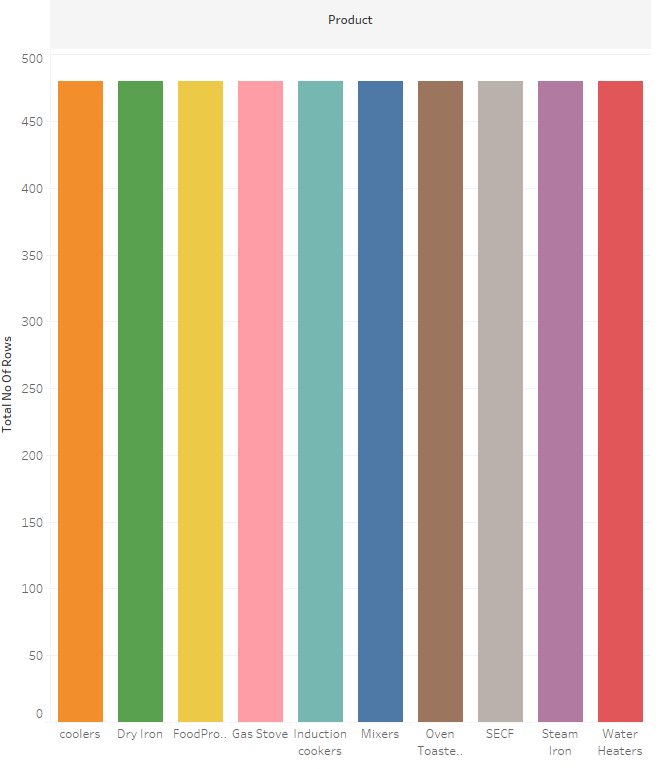
Sales is right skewed and ranging from 0 to 14 M

75% data is between 0 and 2.5 M

## City distribution



## Product Distribution



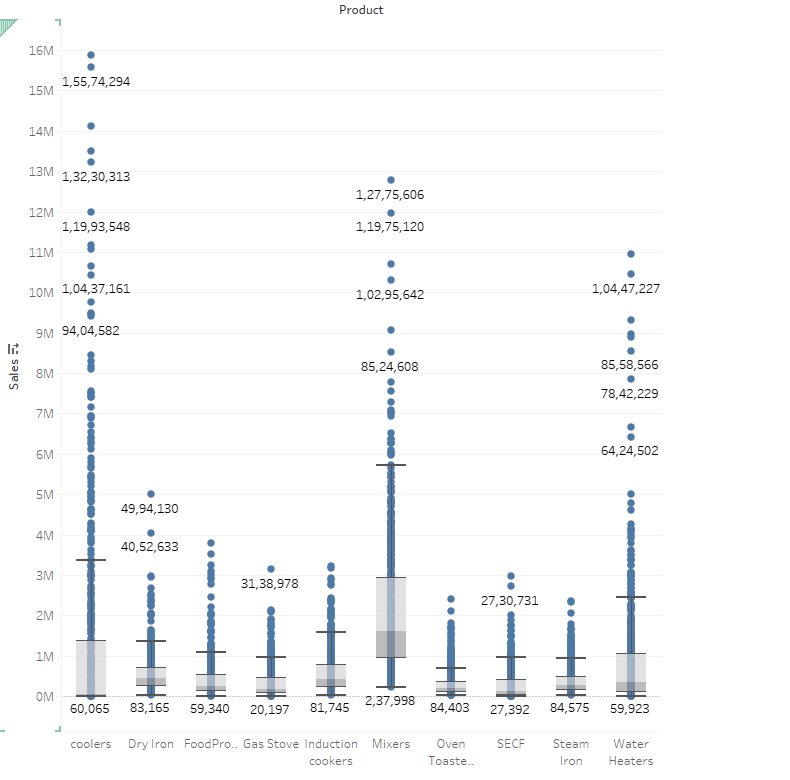
City and product are equally distributed each having 480 rows individually

There are total of 10 different cities in geographical area to be explored

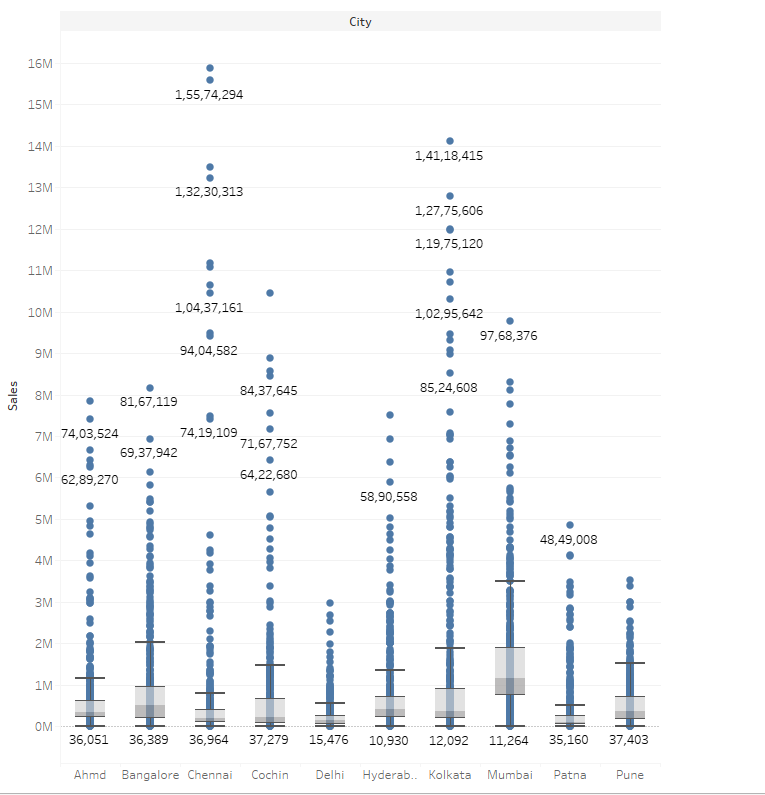
There are 10 different product in different cities, for which sales has to be forecasted

Date range is from April 2014 to March 2018 containing 1st date of every month.

## Box plot for Sales and product



Box plot for Sales and city



All products have outliers

Coolers , Mixers and Water heaters have the highest outliers

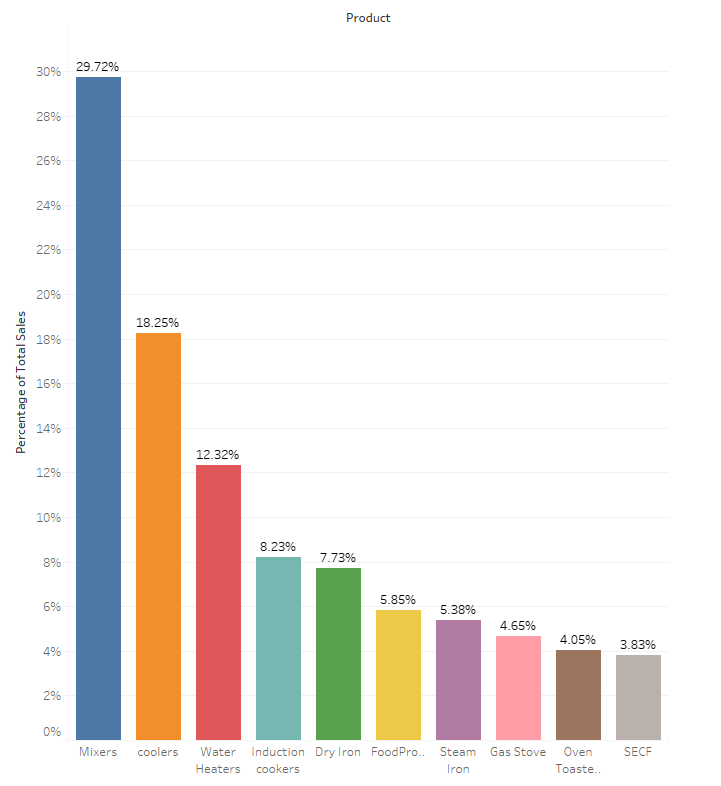
All Cities have outliers

Kolkata, Chennai and cochin have the highest outliers

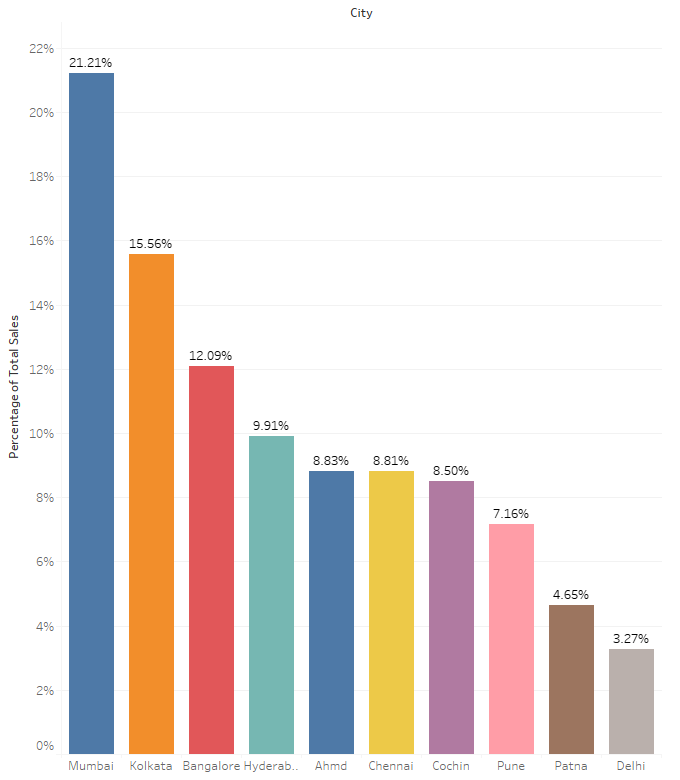
We are not treating outliers because time series forecasting requires continuous data and treating outliers would not give us best outcomes.

## Percentage of Sales and Product

## 



## Percentage of Sales and City



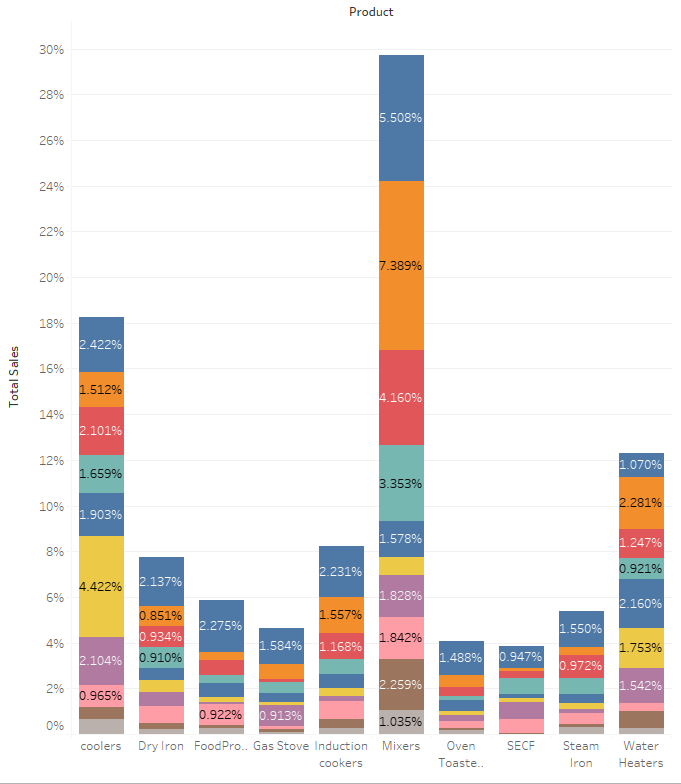
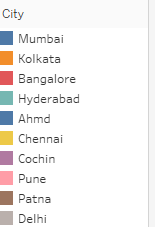
Sales of mixers is the highest and SECF are the lowest

Similarly, Mumbai has highest sales of products and Delhi recorded lowest among others.

Based on above fig, we can concentrate more sales of mixers and adequate numbers of products in stock in Mumbai

For product SECF, production can be minimised based on sales pattern over the years and products sold in Delhi have to be revisited to understand, why sales drops and production of products and sales business strategies to be carried on.

## City and product and sales percentage

Above fig gives a detailed understanding for sales across different cities and products

**Coolers**: Chennai has the highest sales and Delhi has the lowest sales.

Production of coolers should be more in Chennai and business strategies should be explored for very low sales in Delhi.

**Dry Iron :** Sales of dry iron is highest in Mumbai and lowest in Delhi

Production of dry iron should be more in Mumbai and business strategies should be explored for very low sales in Delhi.

**Food processor :** Sales are high in Mumbai and Lowest in Delhi

Production of Food processor should be more in Mumbai and business strategies should be explored for very low sales in Cochin.

**Gas Stove :** Sales are high in Mumbai and Lowest in Delhi

Production of Gas Stove should be more in Mumbai and business strategies should be explored for very low sales in Delhi.

**Induction Cookers** : Sales are high in Mumbai and Lowest in Delhi

Production of Induction Cookers should be more in Mumbai and business strategies should be explored for very low sales in Delhi.

**Mixers :** Sales are high in Kolkata and Lowest in Delhi

Production of Mixers should be more in Kolkata and business strategies should be explored for very low sales in Delhi.

**Oven Toaster grill :** Sales are high in Mumbai and Lowest in Delhi

Production of Oven toaster grill should be more in Mumbai and business strategies should be explored for very low sales in Delhi.

**SECF :** Sales are high in Mumbai and Lowest in Delhi

Production of SECF should be more in Mumbai and business strategies should be explored for very low sales in Delhi.

**Steam Iron :** Sales are high in Mumbai and Lowest in Delhi

Production of Steam Iron should be more in Mumbai and business strategies should be explored for very low sales in Delhi.

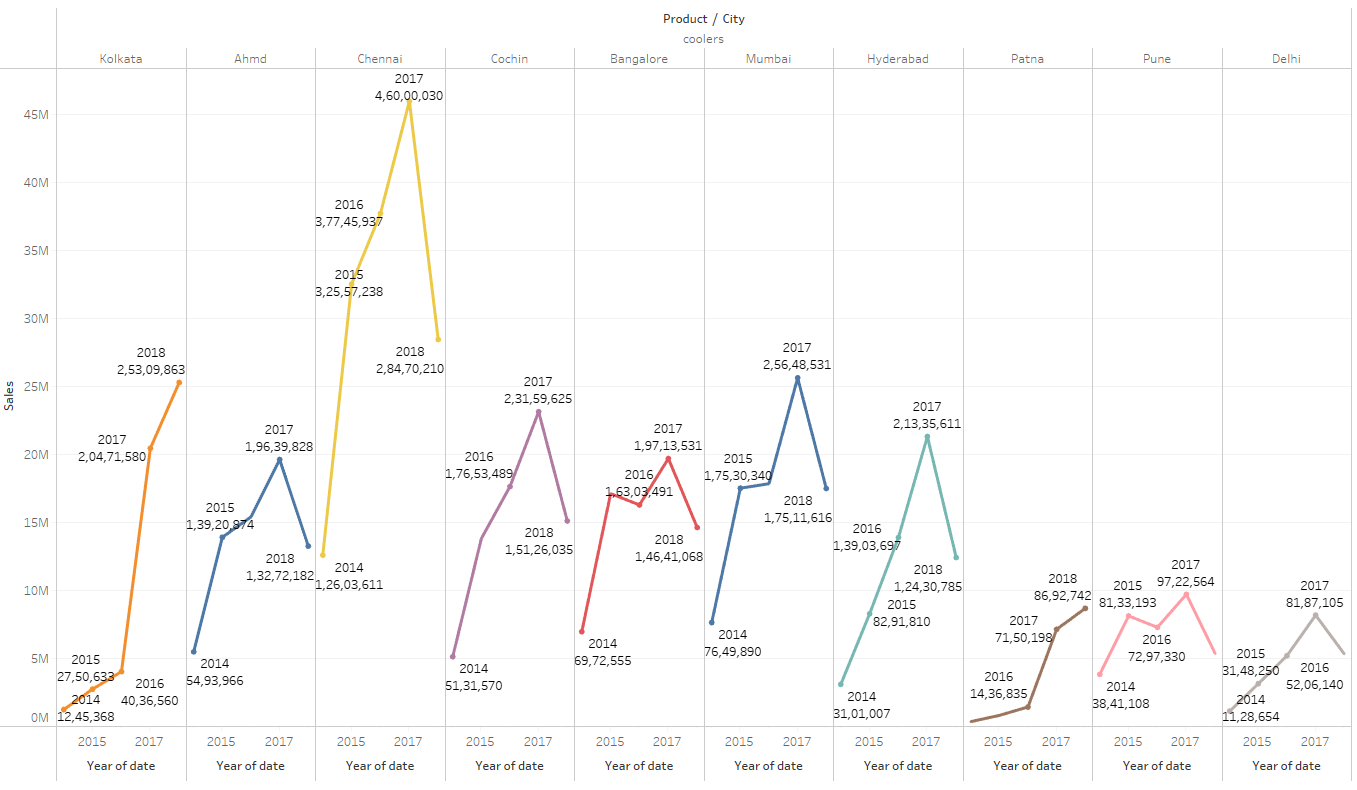
**Water Heater :** Sales are high in Kolkata and Lowest in Delhi

Production of Water Heater should be more in Kolkata and business strategies should be explored for very ow sales in Delhi.

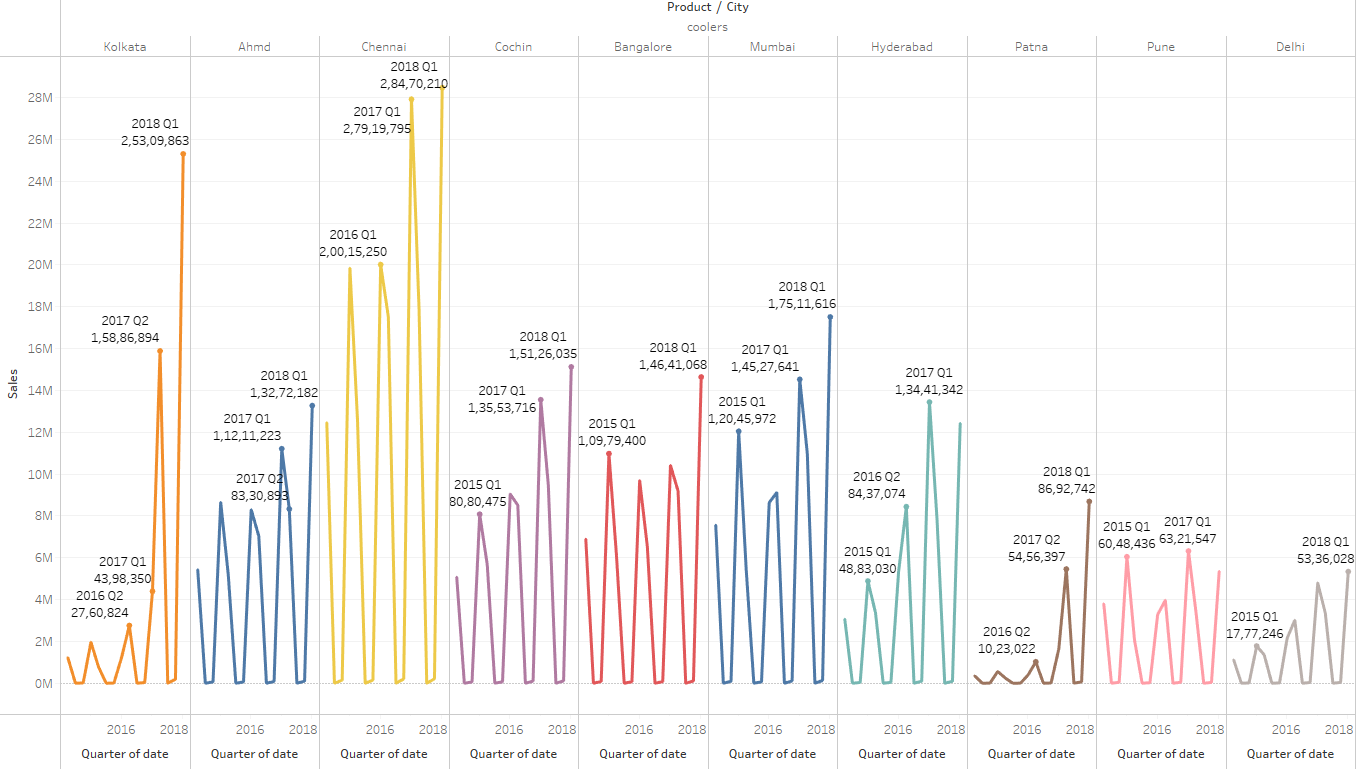
# Date and city and sales and product

## Cooler vs Cities vs Date

Yearly



Quarterly



Monthly



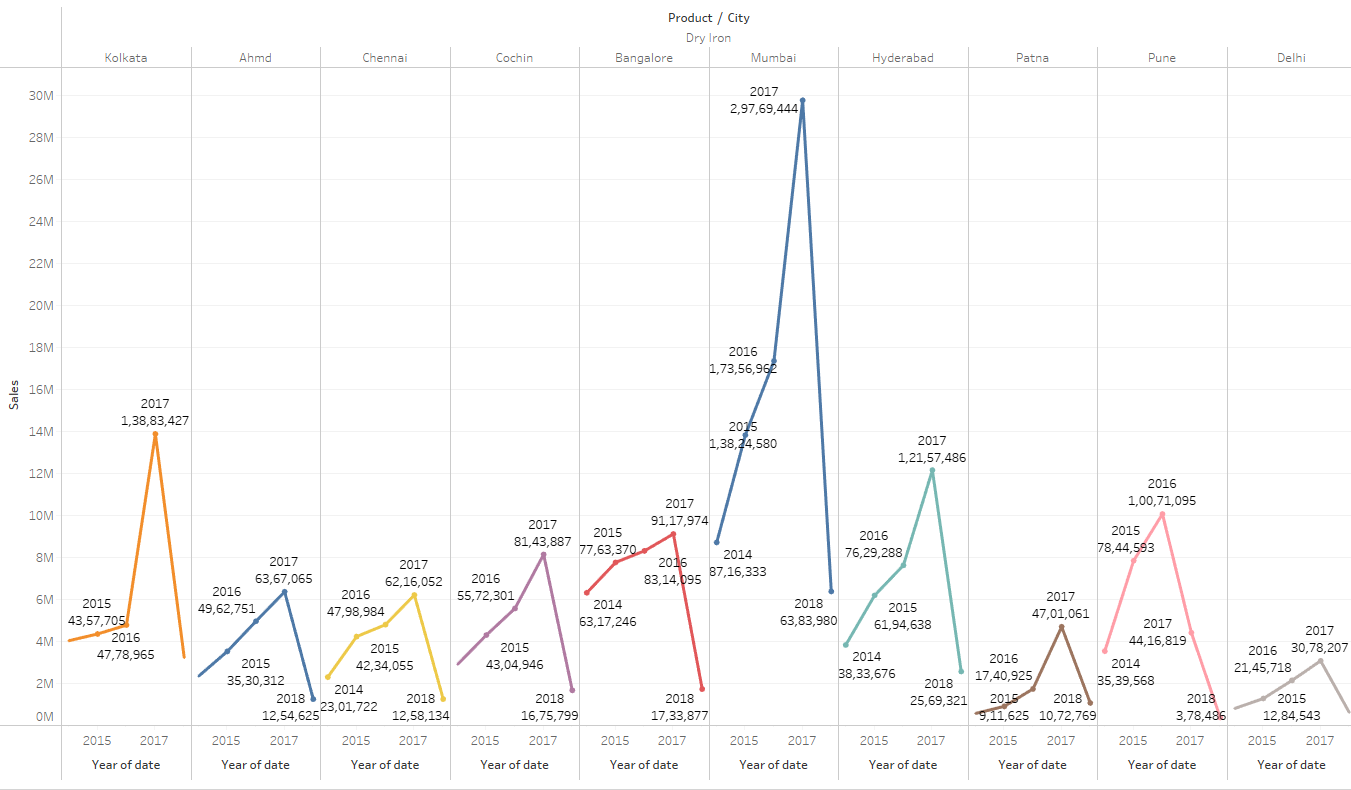
Yearly – Kolkata, Chennai, Patna, Delhi and Cochin shows an increase in sales steadily from 2014 to 2017 and dip in Q1 of 2018. Ahmedabad, Bangalore, Mumbai, Pune increases in sales from 2014, but there is a sudden decreases in sales in between and gradual increase till 2017.

Quarterly - There is a pattern of decrease in sales during Q3 and Q4 of Cooler across all cities

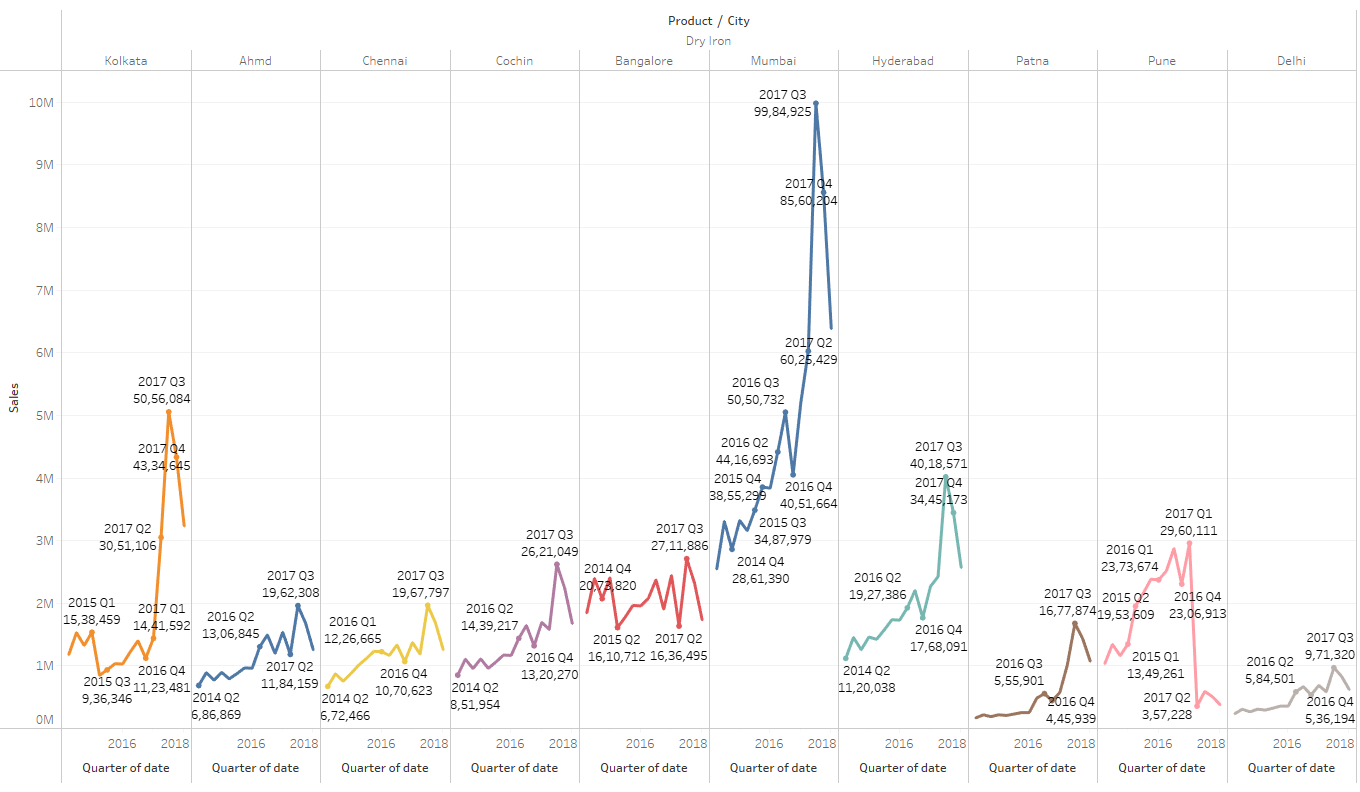
Monthly – July to December there is decrease in sales and march has the highest sales across all cities. There is increase in trend line in all cities except in Pune

## Dry Iron vs Cities vs Date

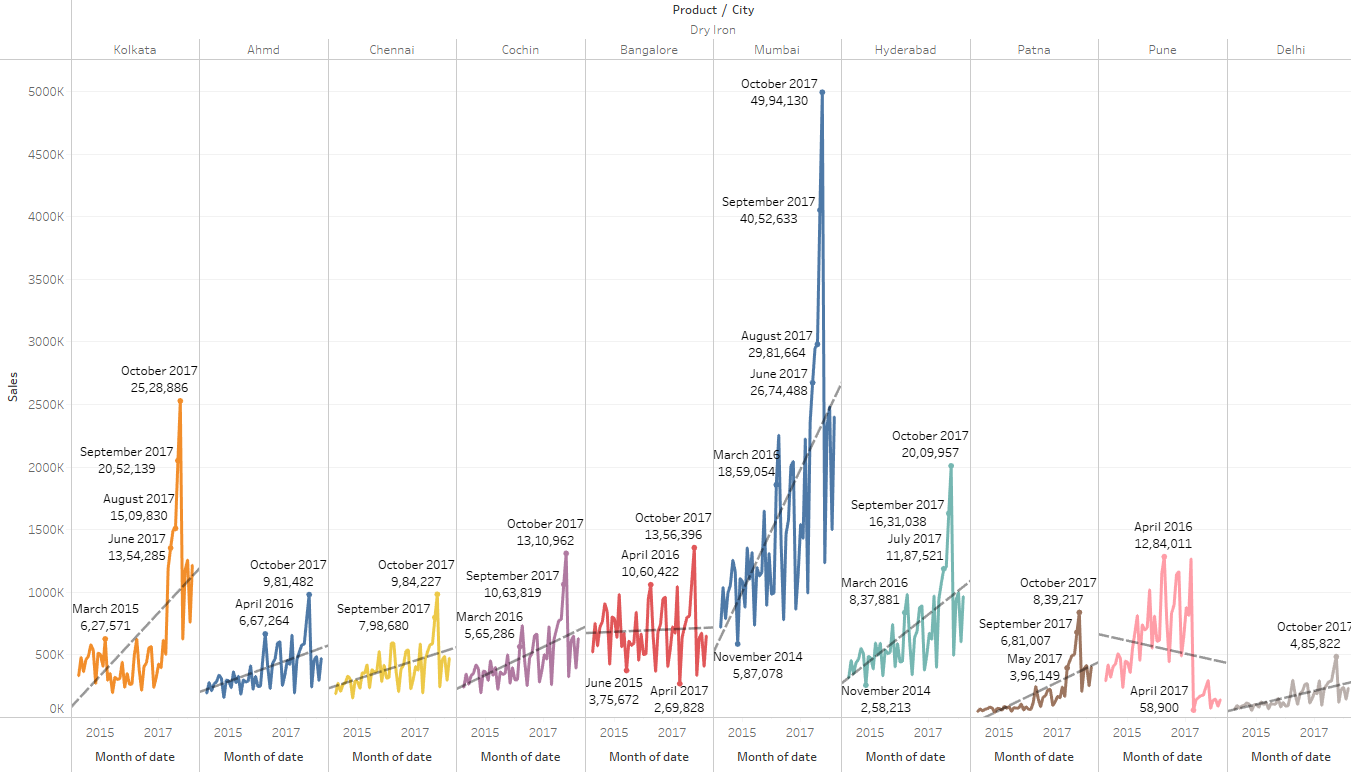
Yearly



Quarterly



Monthly



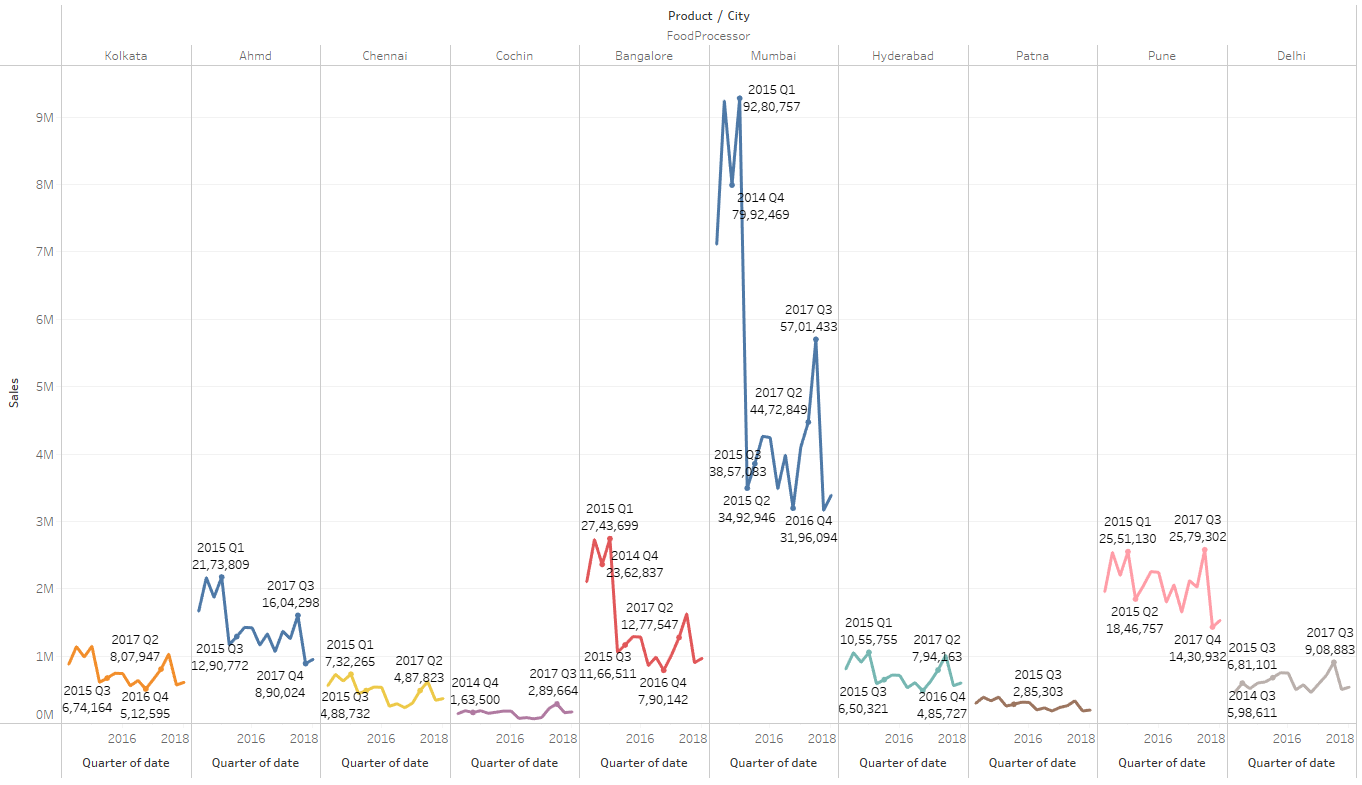
Yearly – All cities shows an increase in sales steadily from 2014 to 2017 and dip in Q1 of 2018. Mumbai shows promising sales than other cities

Quarterly - No pattern of increase or decrease in sales. But sales gradually increases with slight dip in certain months

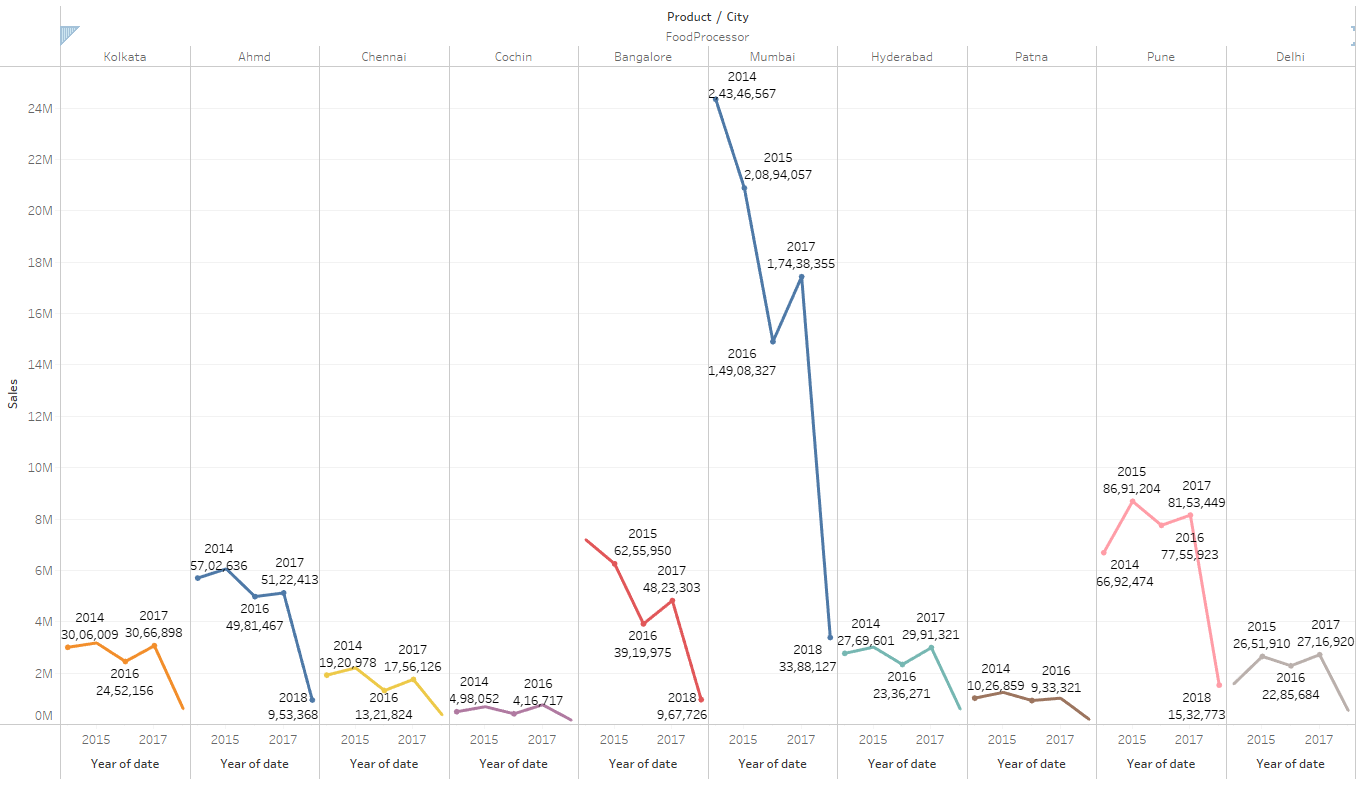
Monthly –October has the highest sales across all cities. There is increase in trend line in all cities except in Pune

## Food Processor vs Cities vs Date

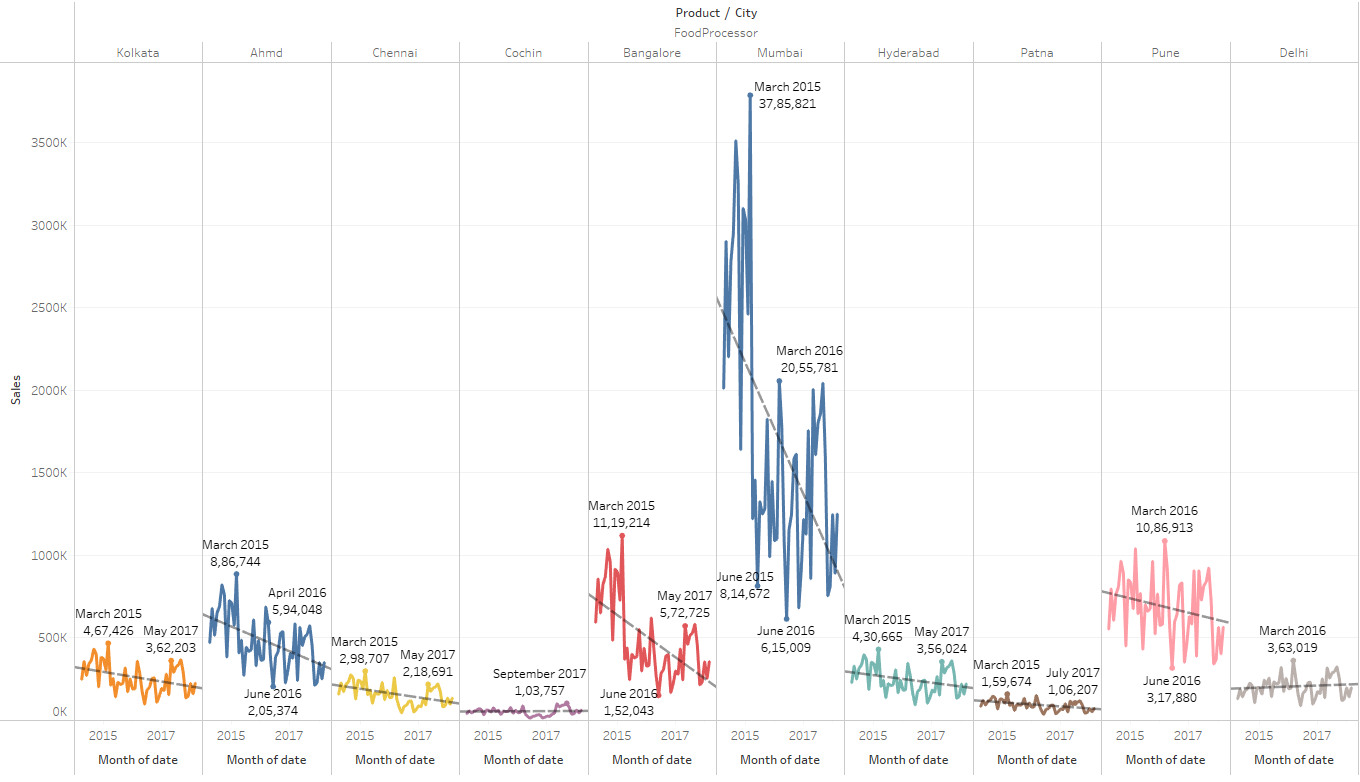
Yearly



Quarterly



Yearly



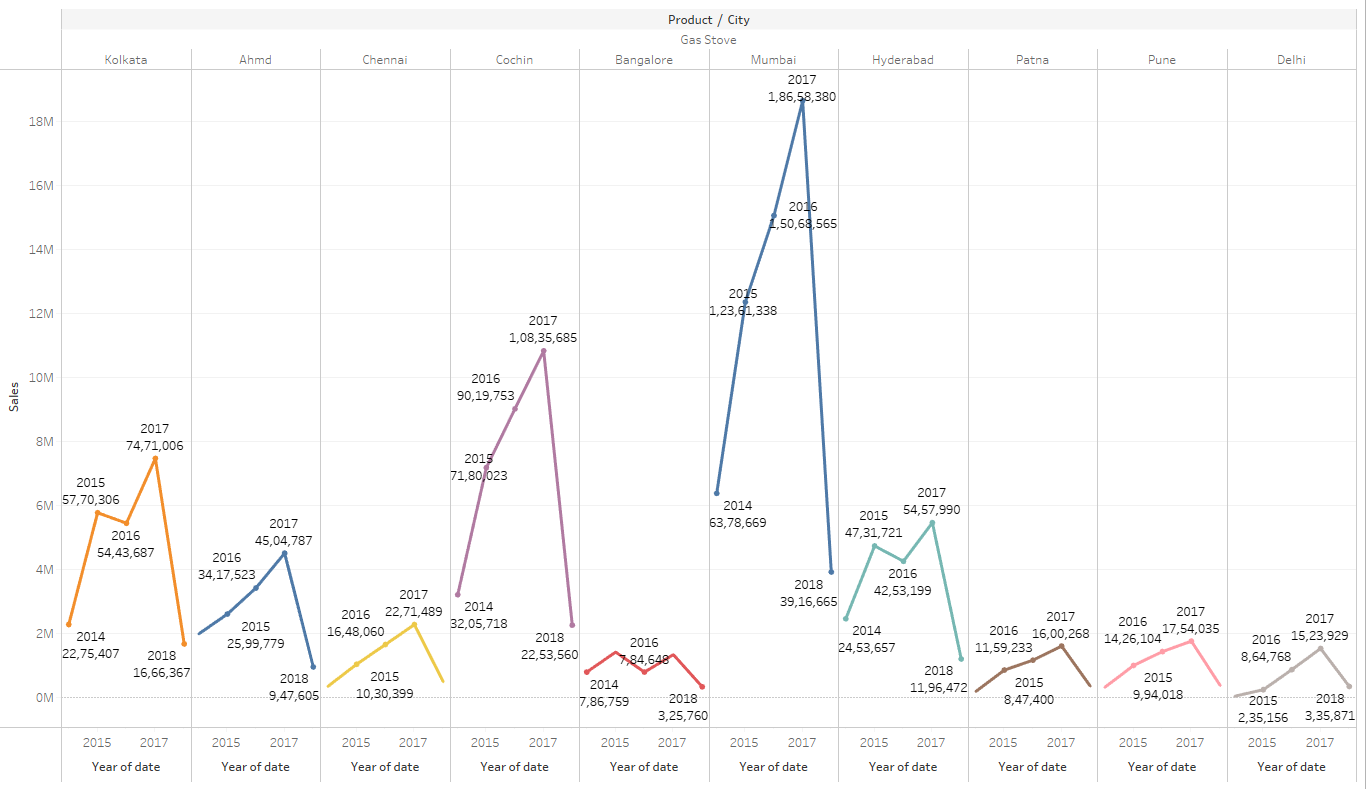
Yearly – All cities shows decrease in sales steadily from 2014 to 2017 and slight increase in sales in 2017, 2015 of 2014. Mumbai had highest sales during 2014 than other cities

Quarterly - No pattern of increase or decrease in sales. But sales gradually decreases

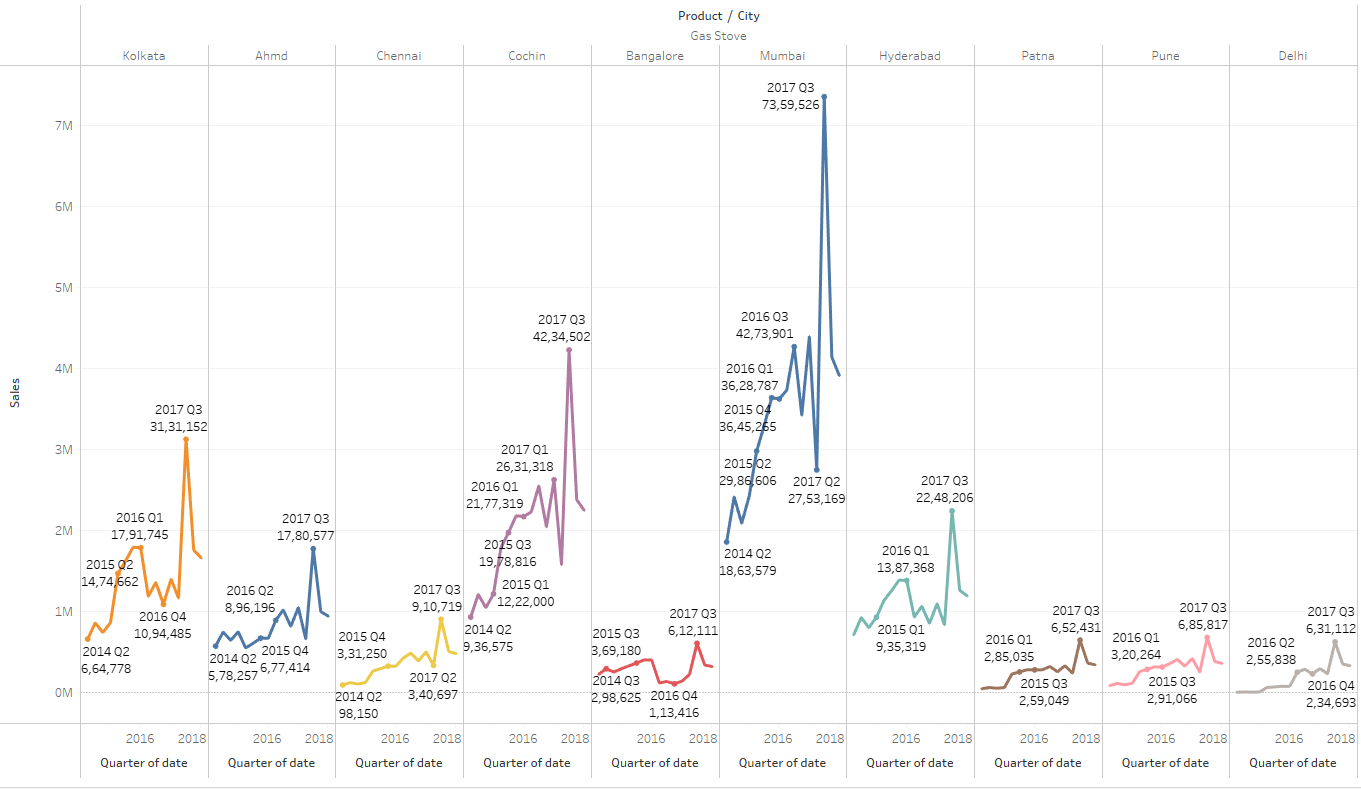
Monthly –Mostly march has the highest sales across all cities except Cochin. There is decrease in trend line in all cities except Delhi, which has flat trend line

Gas Stove vs Cities vs Date

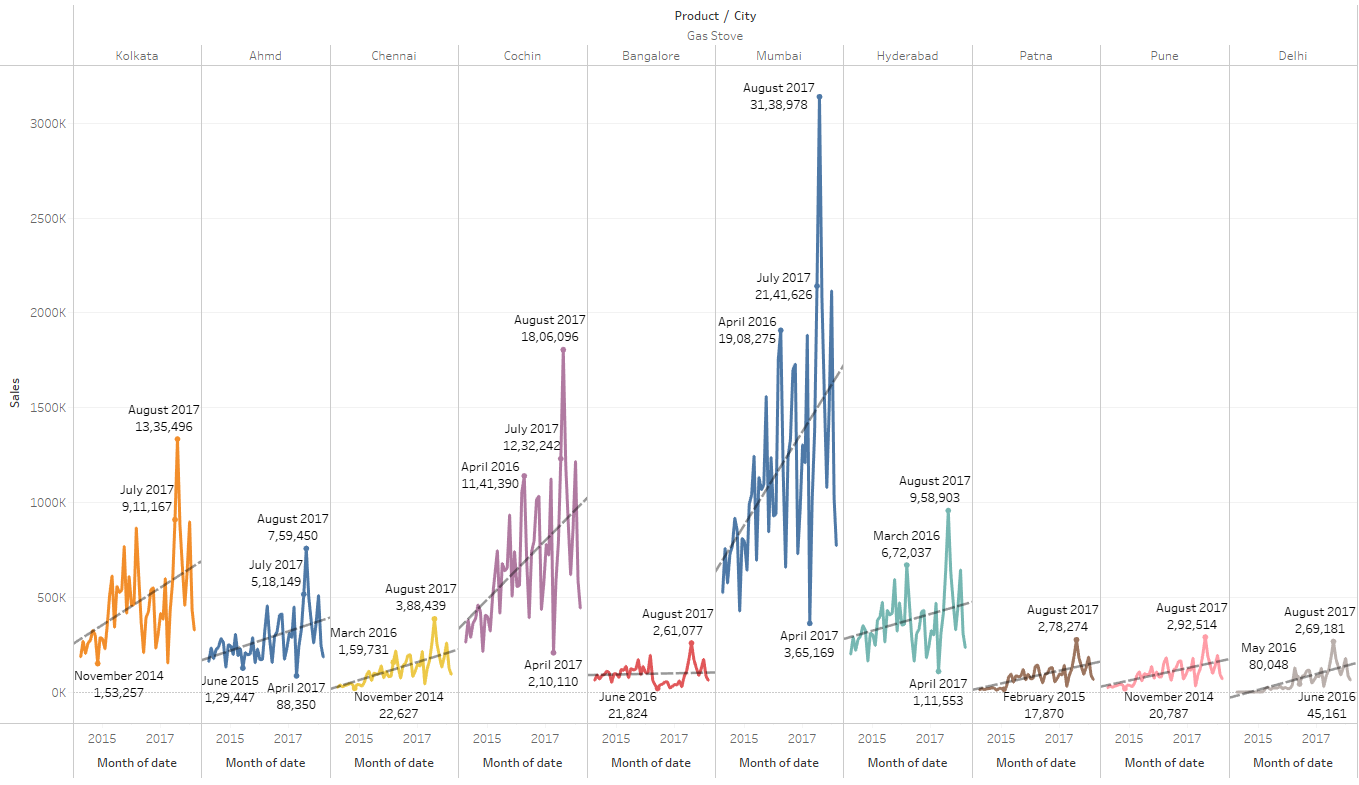
Yearly



Quarterly



Monthly

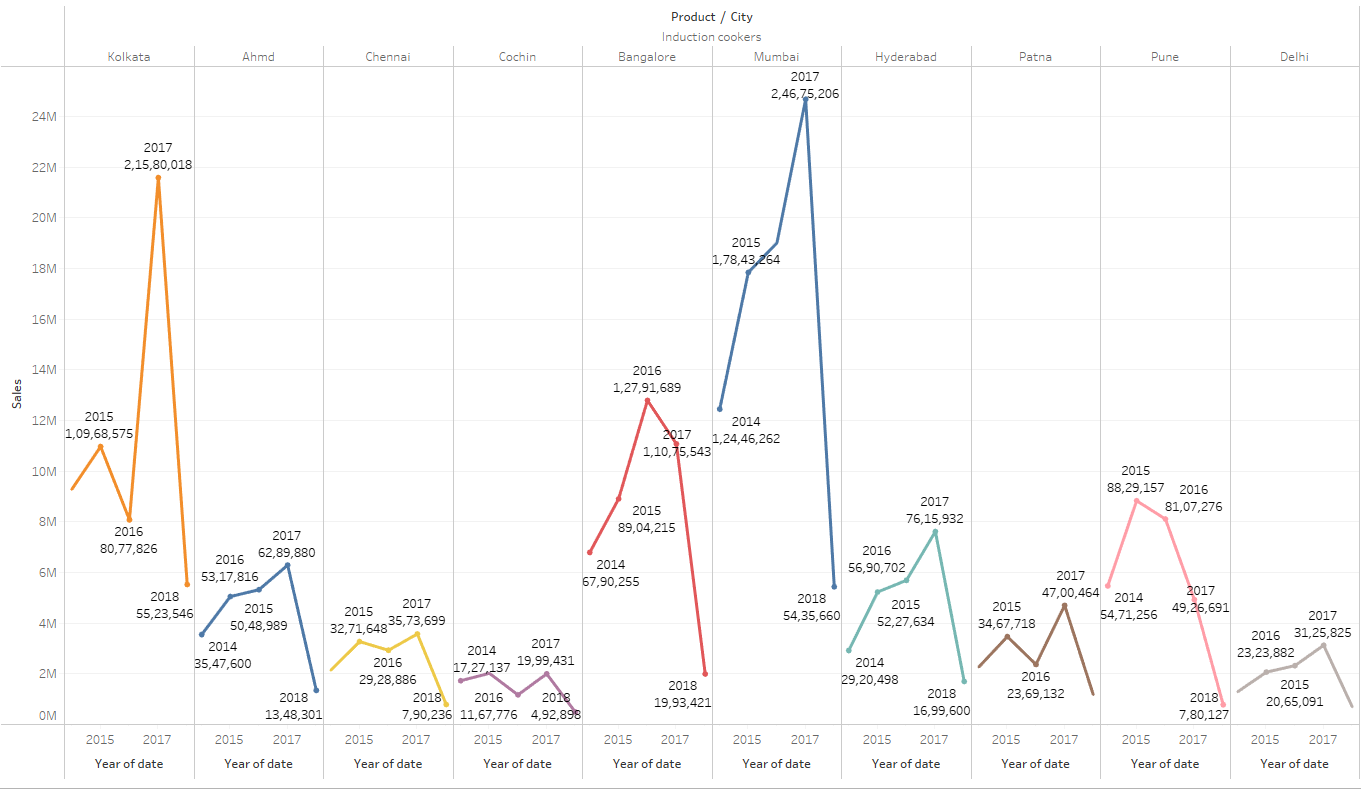


Yearly – Ahmedabad, Chennai, Cochin, Mumbai, Patna, Pune and Delhi shows an increase in sales steadily from 2014 to 2017 and dip in Q1 of 2018. Whereas Kolkata, Bangalore, Hyderabad have dip in sales in between.

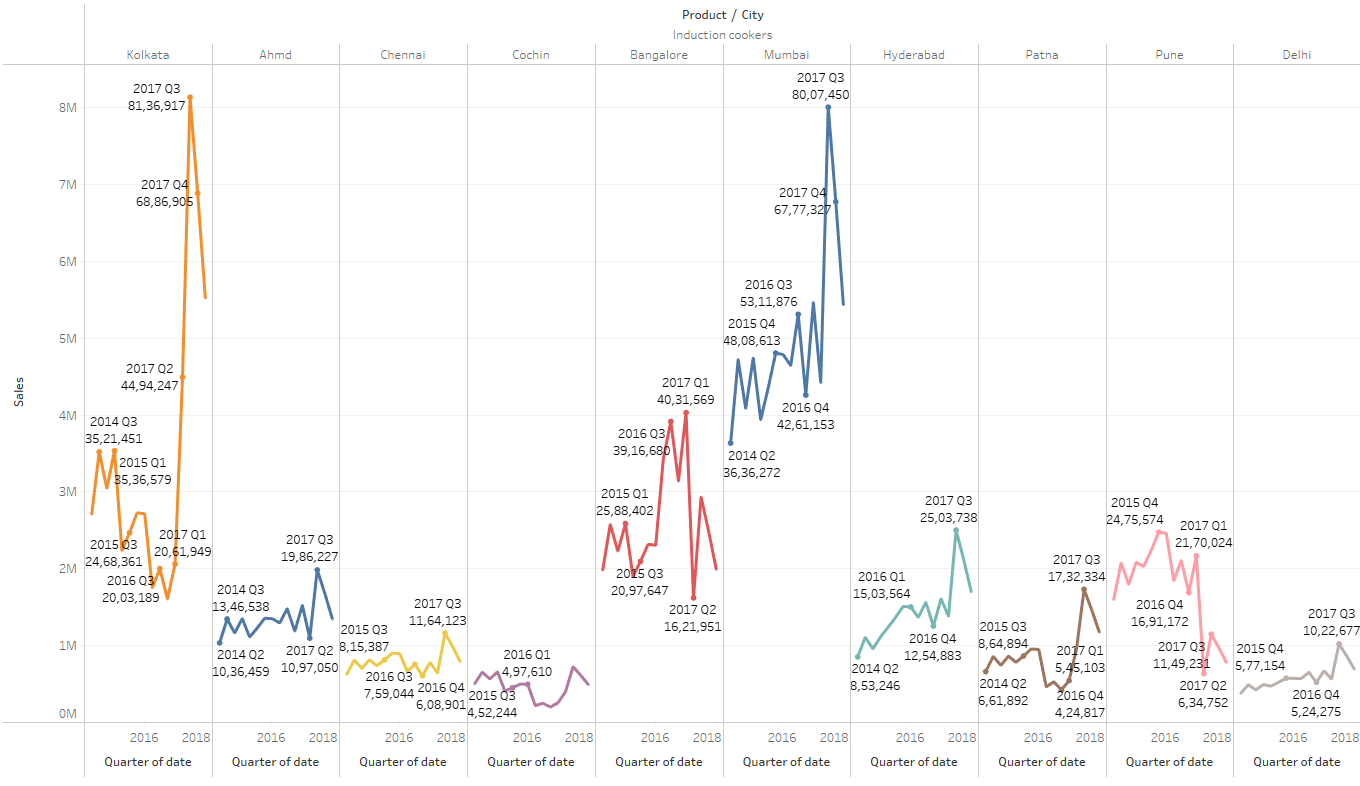
Quarterly - Q3 has the highest number of sales across all cities. But sales gradually increases with slight dip in certain months

Monthly –August has the highest sales across all cities. There is increase in trend line in all cities except in Bangalore , which has a flat line.

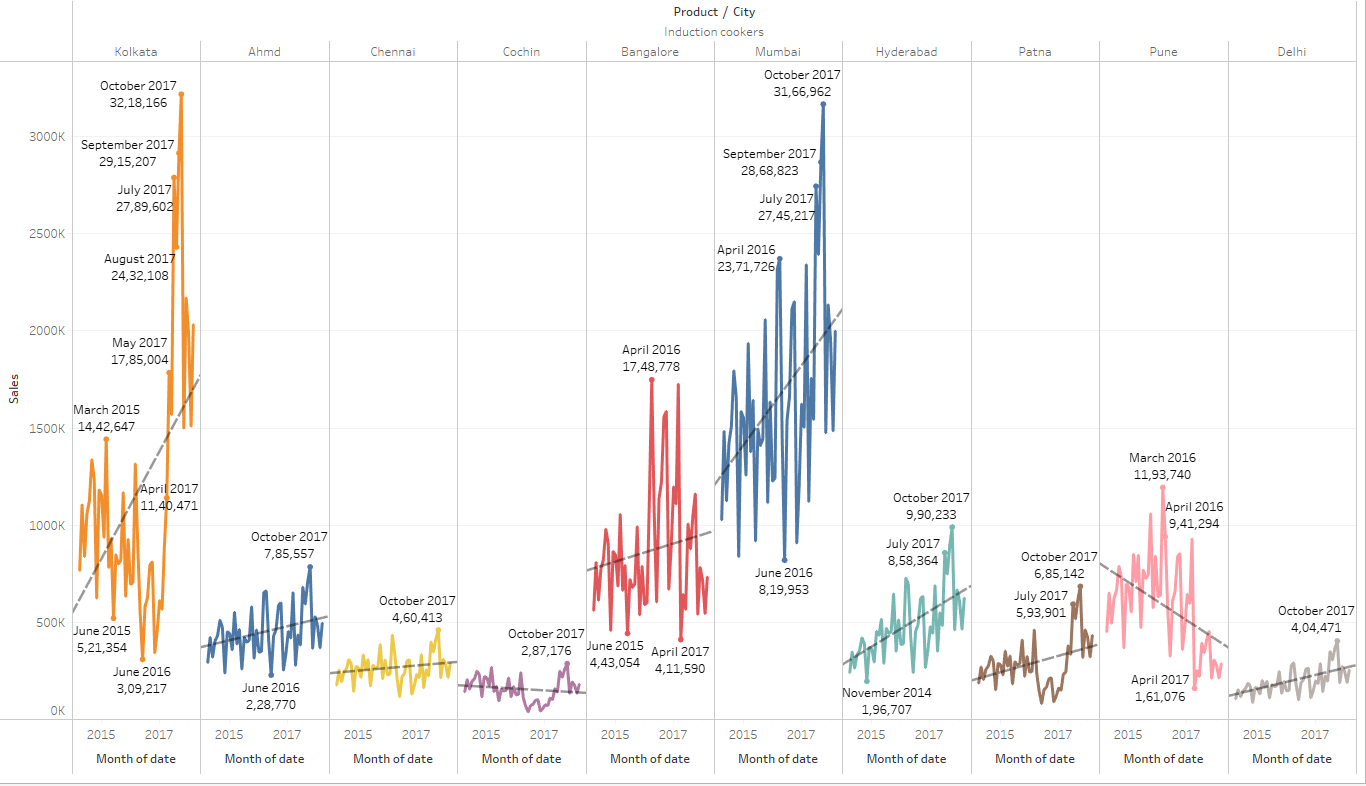
Induction Cookers vs Cities vs Date

Yearly

Quarterly



Yearly



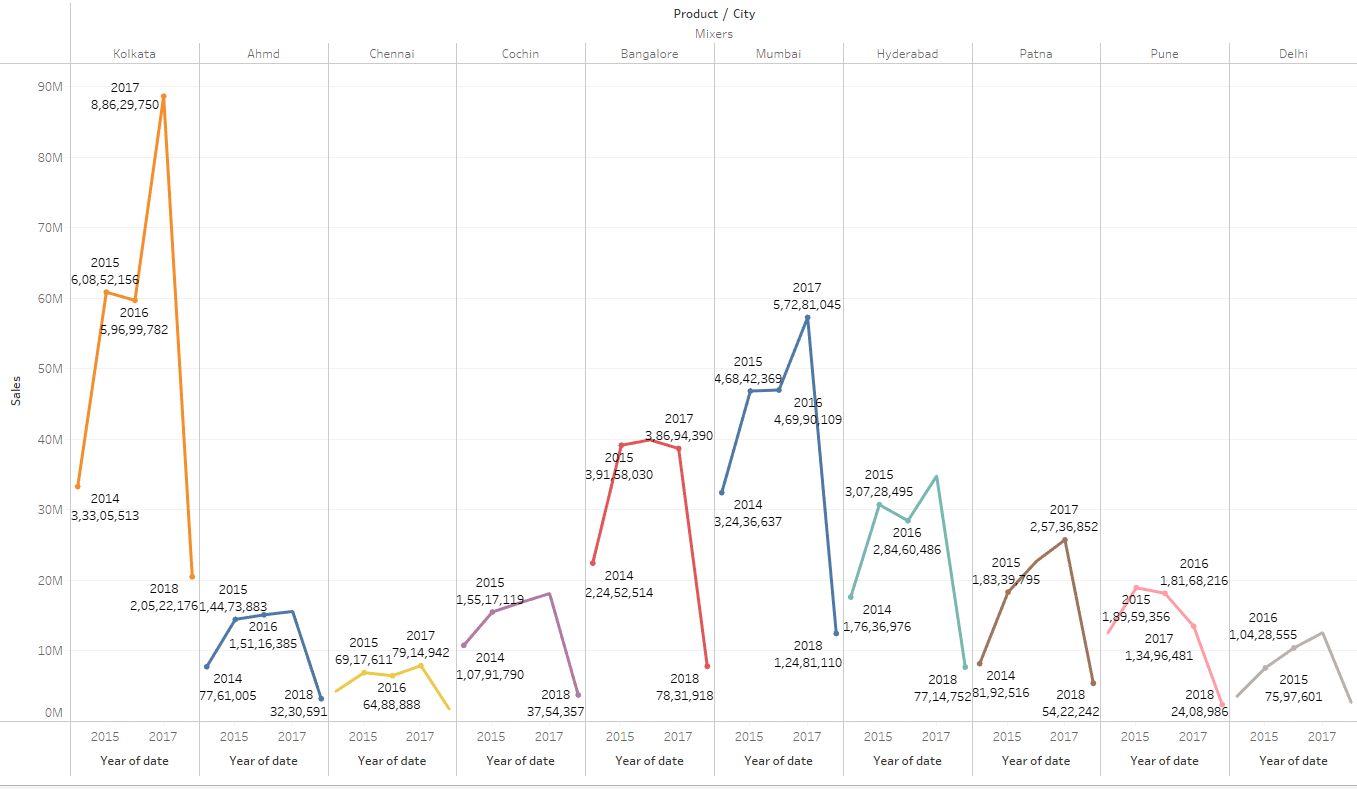
Yearly – Mumbai shows promising sales than other cities. Sales of induction cookers is not consistent in all cities and it varies across years

Quarterly - No pattern of increase or decrease in sales. But sales gradually increases with slight dip in certain months except Pune, where sales decreases from 2017 Q1

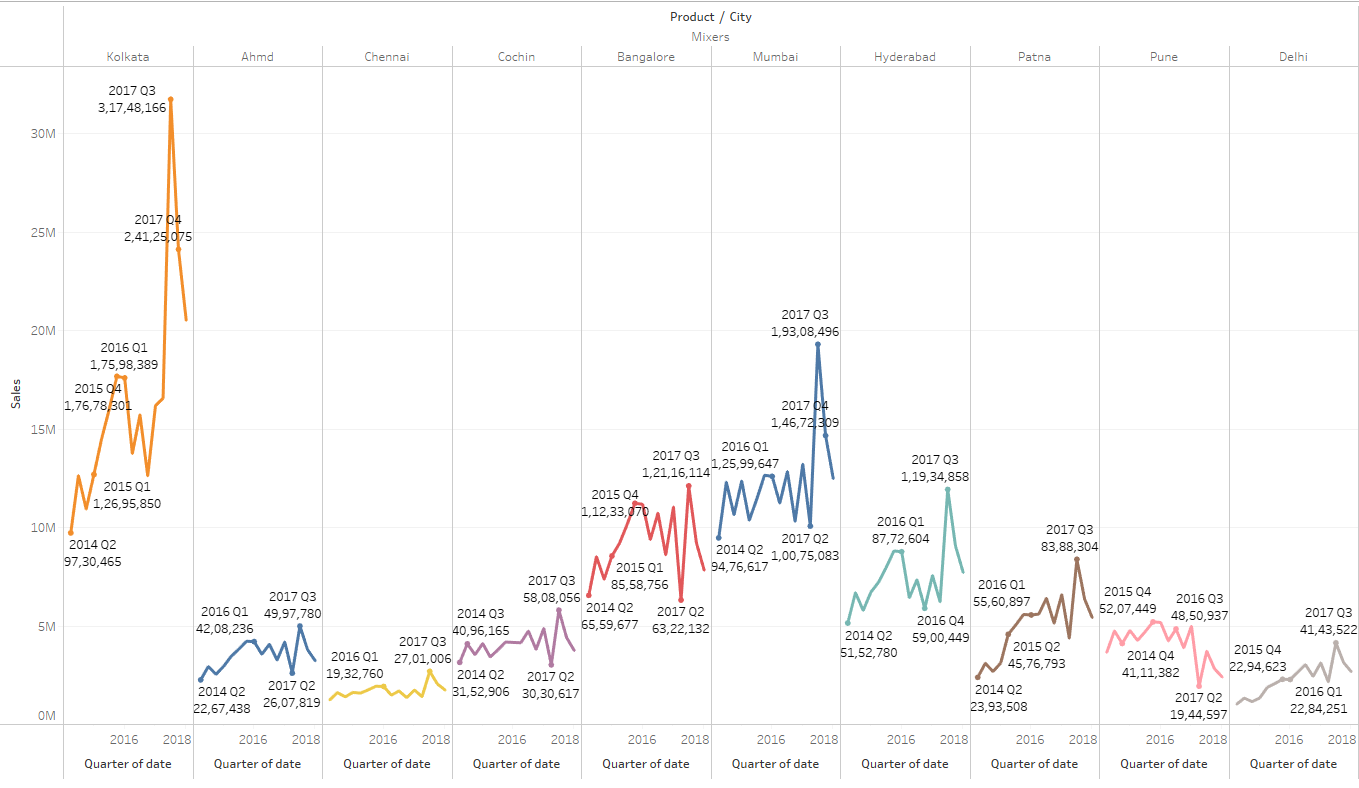
Monthly –October has the highest sales across all cities except Pune and Bangalore. There is increase in trend line in all cities except in Pune

Mixers vs Cities vs date

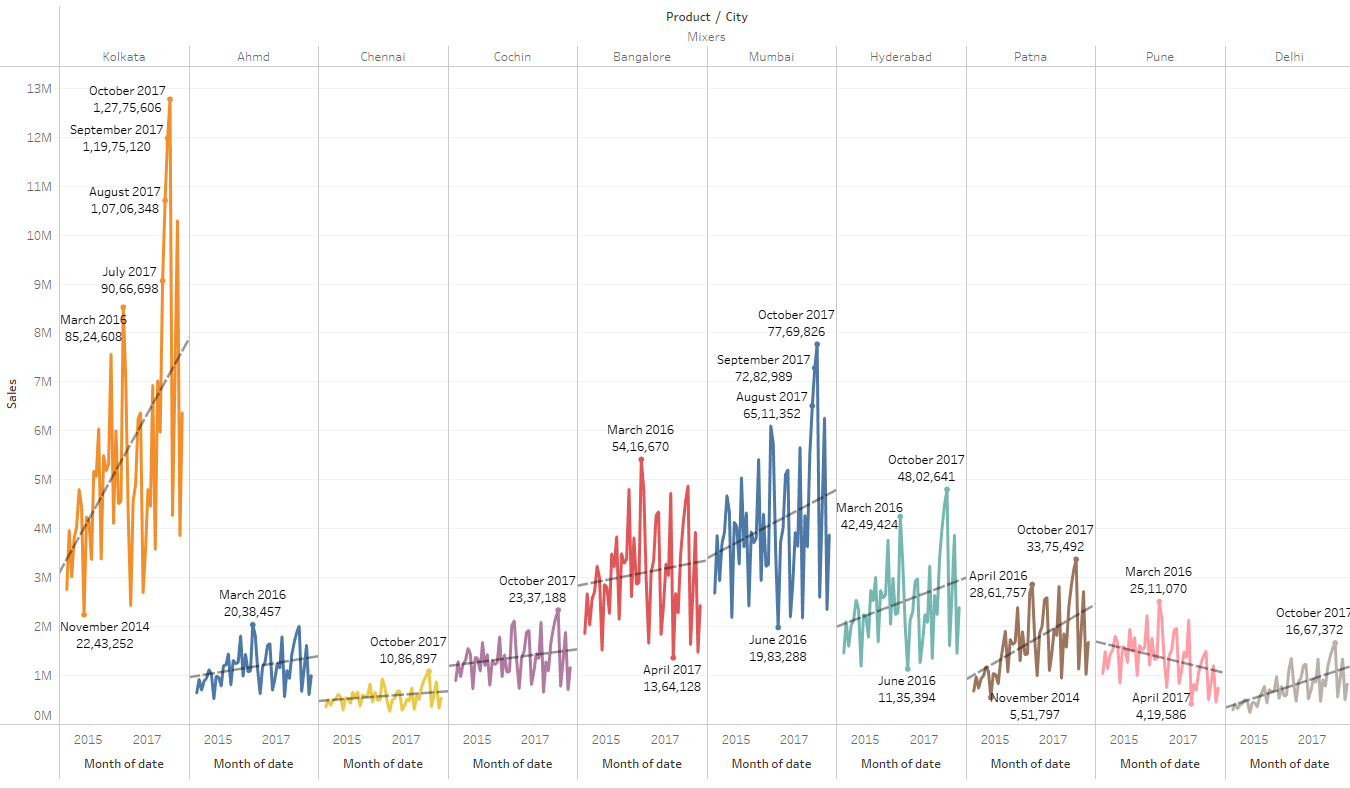
Yearly



Quarterly



Yearly



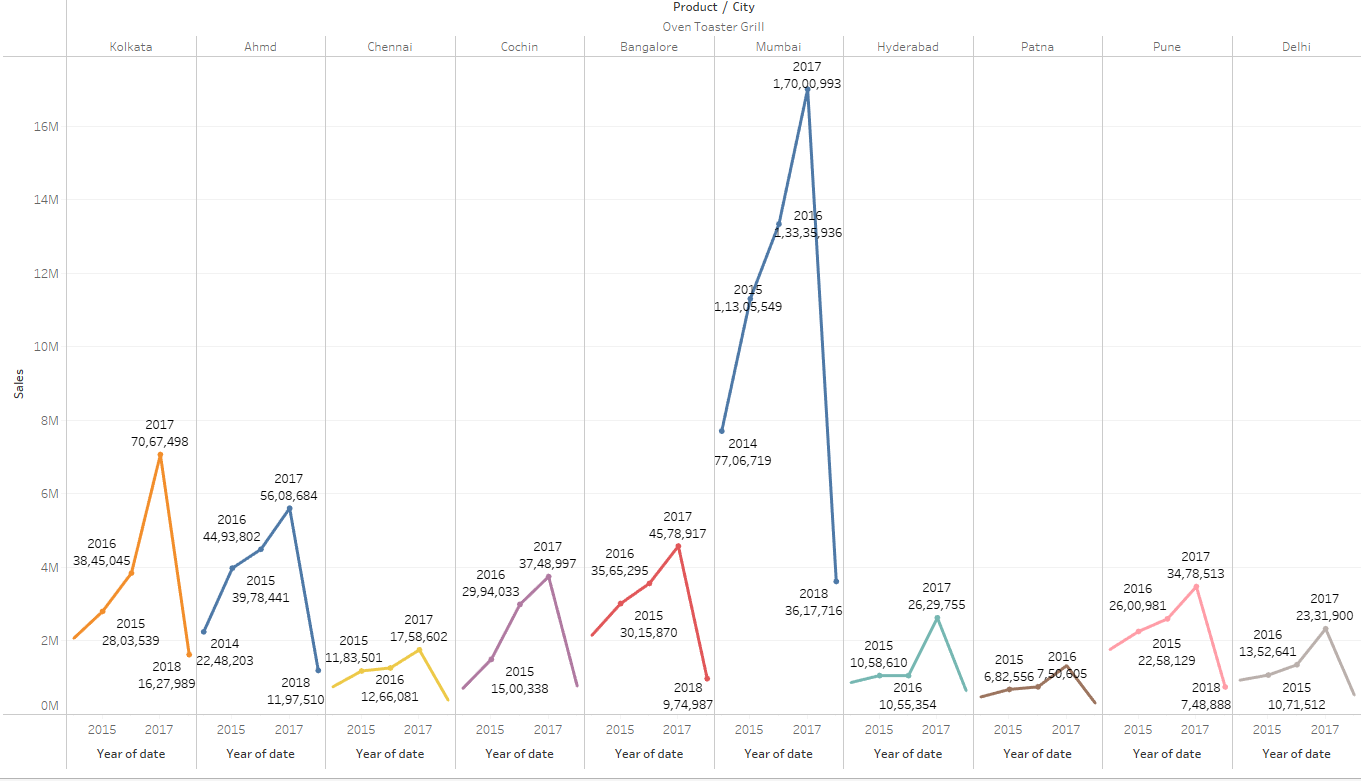
Yearly – Kolkata shows promising sales than other cities. Sales of Mixers is not consistent in all cities and it varies across all years.

Quarterly - No pattern of increase or decrease in sales. But sales gradually increases with slight dip in certain months except pune

Monthly –October has the highest sales across all cities except Pune and bangalore. There is increase in trend line in all cities except in Pune

## Oven toaster grill vs Cities vs Date

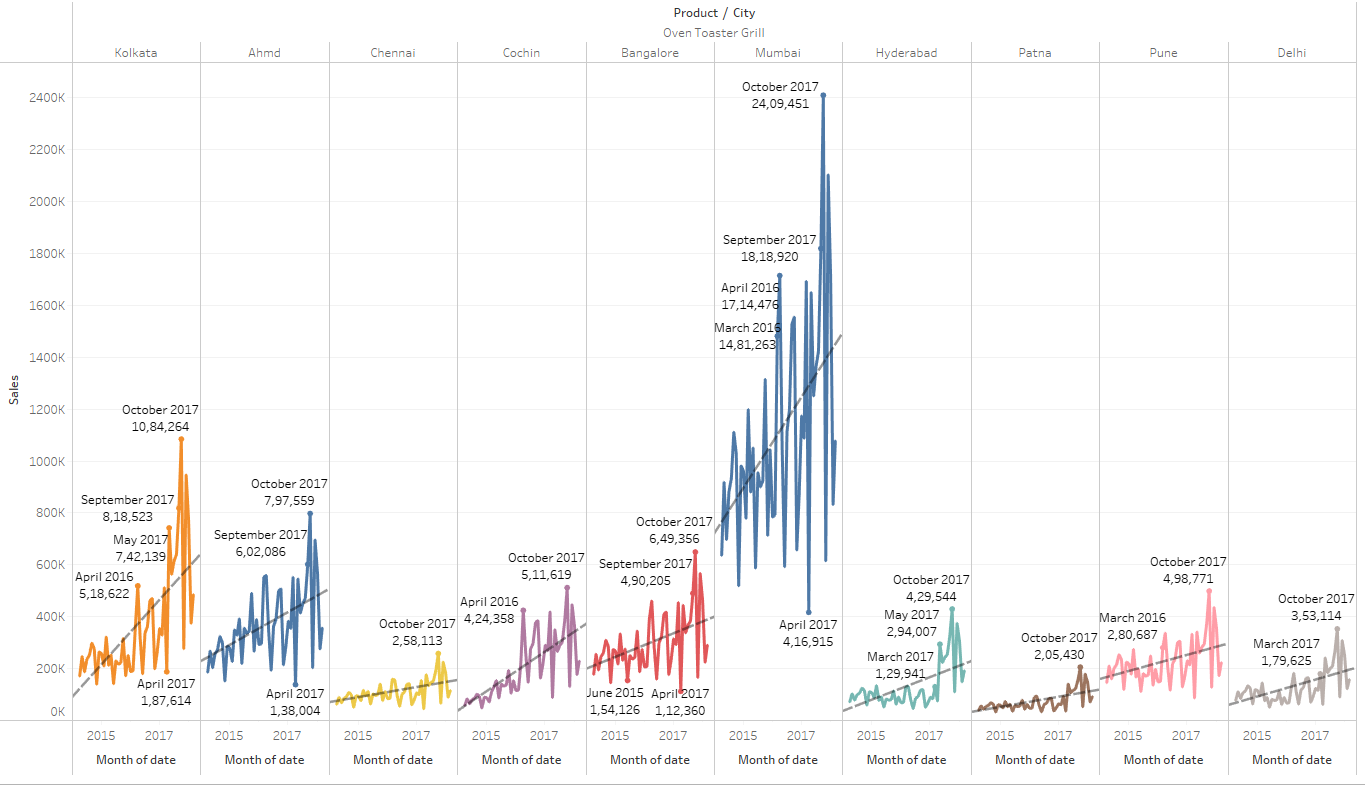
Yearly



Quarterly



Monthly



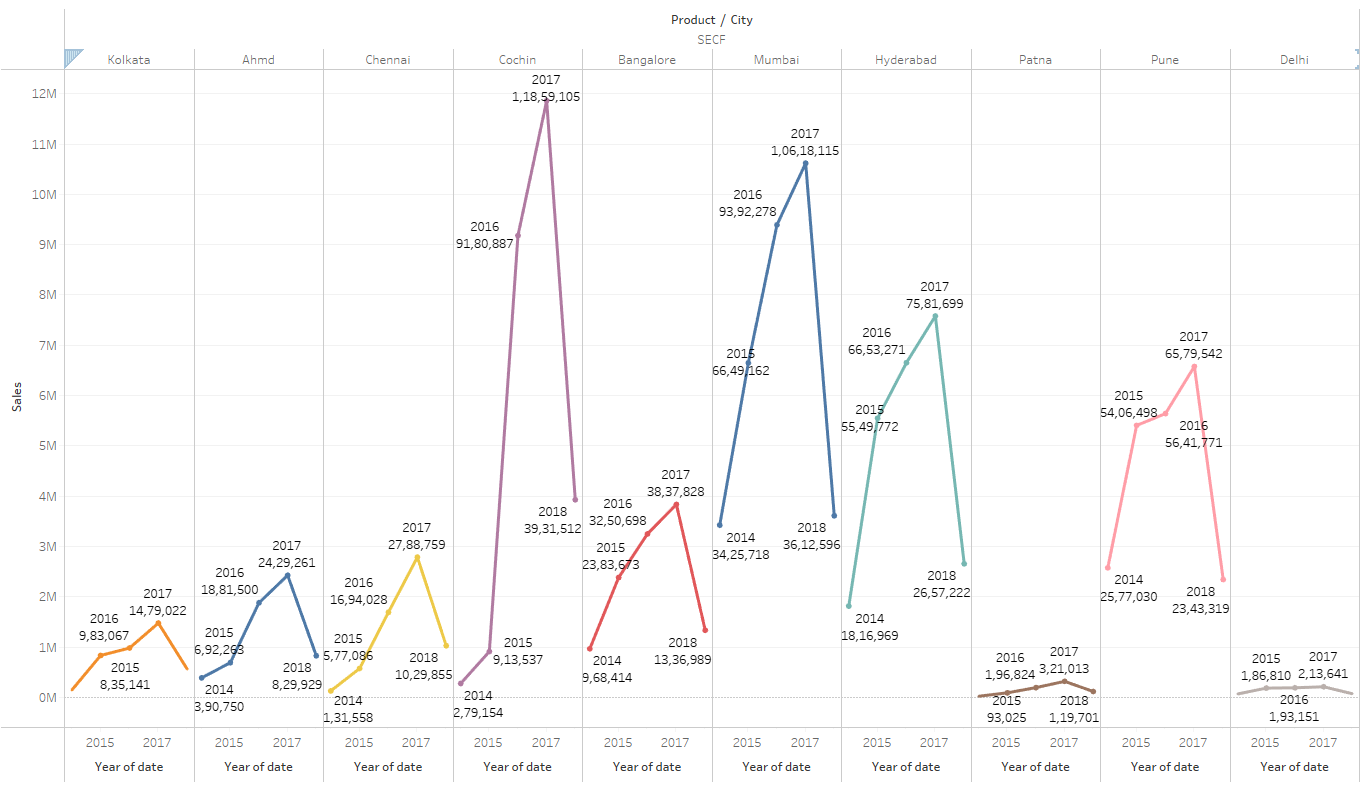
Yearly – Mumbai shows promising sales than other cities. Sales of Oven Toasters gradually increases across years.

Quarterly - No pattern of increase or decrease in sales. But sales gradually increases with slight dip in certain months

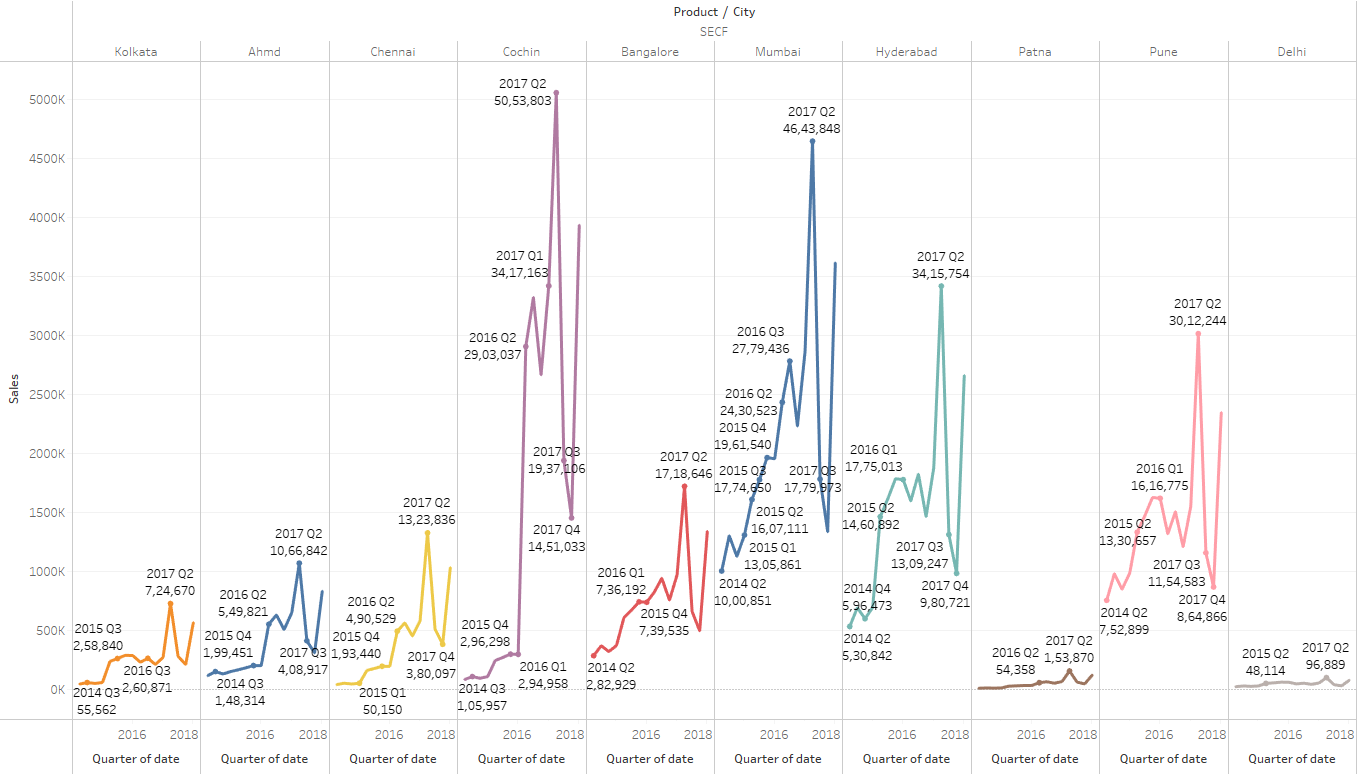
Monthly –October has the highest sales across all cities except Pune and Bangalore. There is increase in trend line in all cities

## SECF vs Cities vs Date

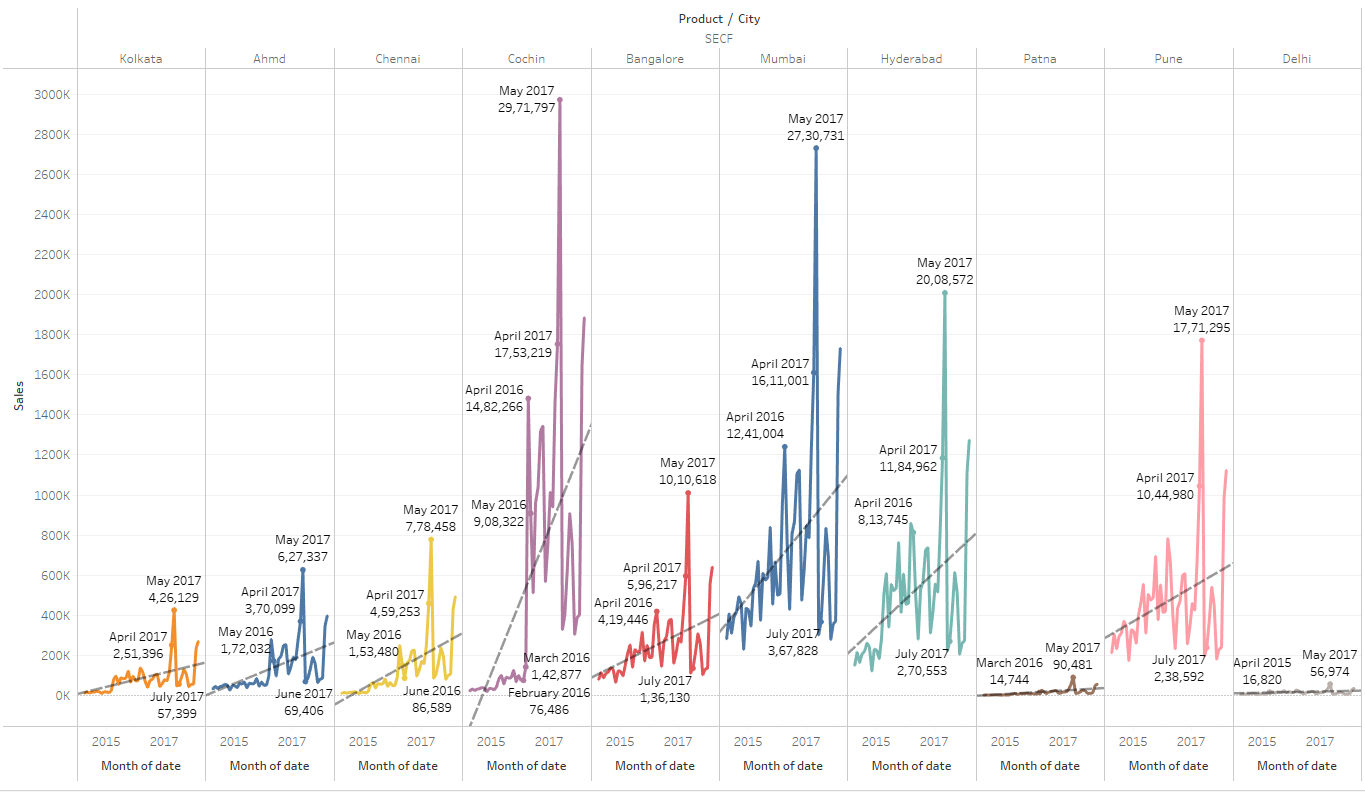
Yearly



Quarterly



Monthly



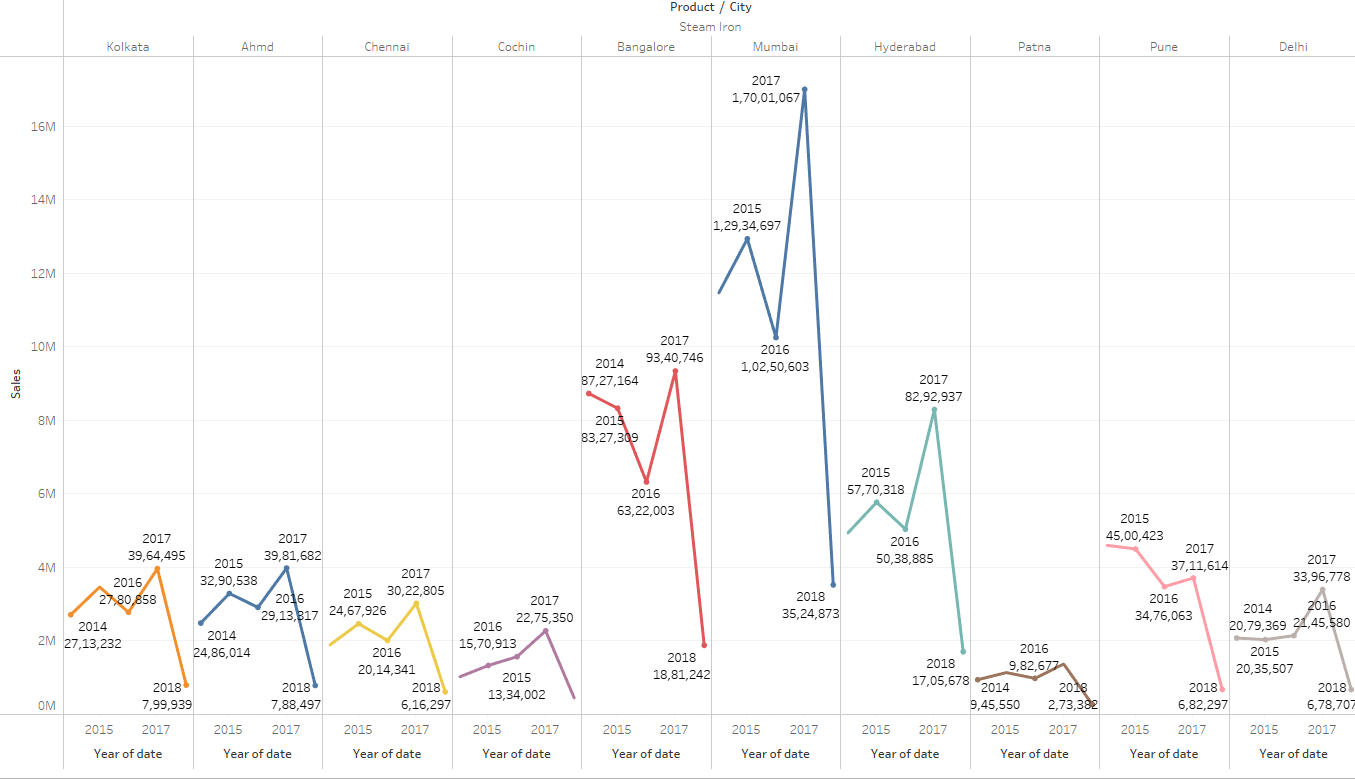
Yearly – Kolkata shows promising sales than other cities. Sales of SECF gradually increases across years.

Quarterly - No pattern of increase or decrease in sales. But sales gradually increases with slight dip in certain months. Q2 shows highest sales across all cities

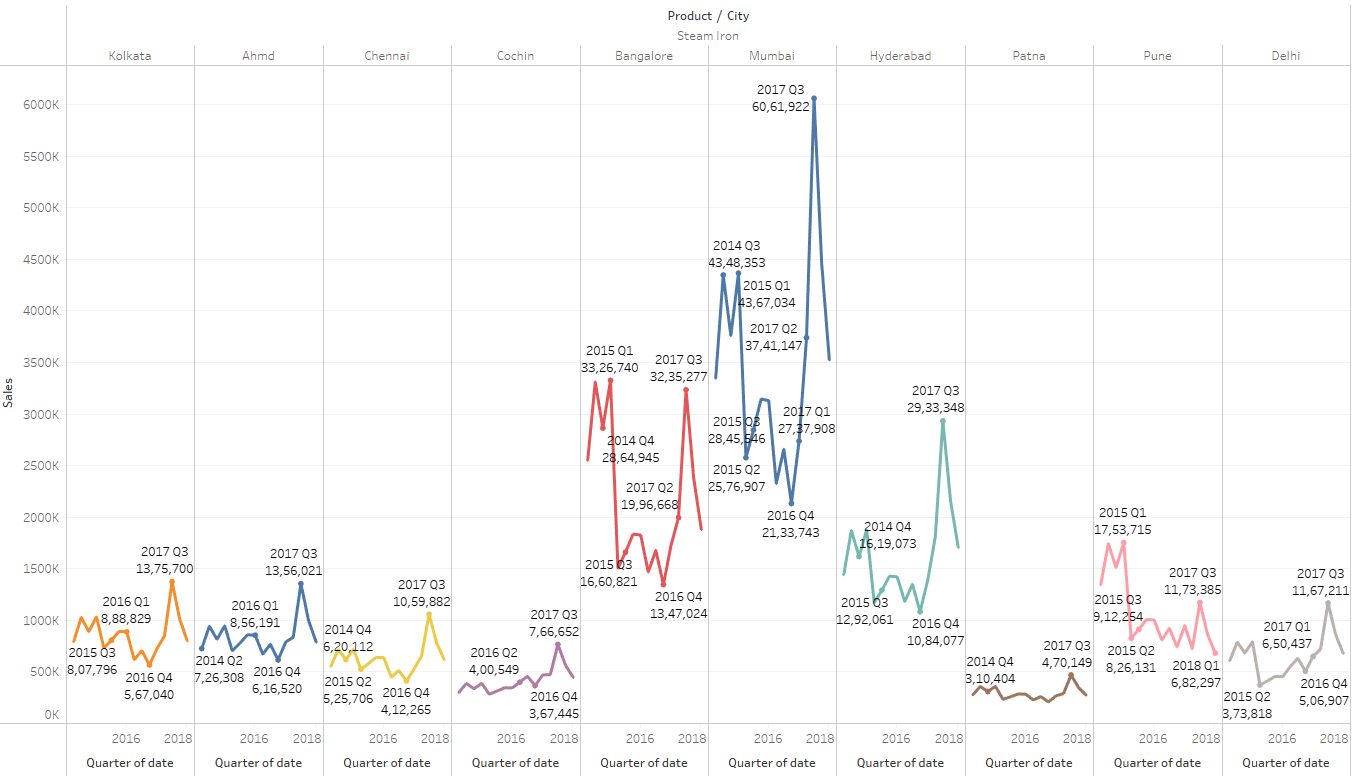
Monthly –May has the highest sales across all cities. There is increase in trend line in all cities

## Steam Iron vs Cities vs Date

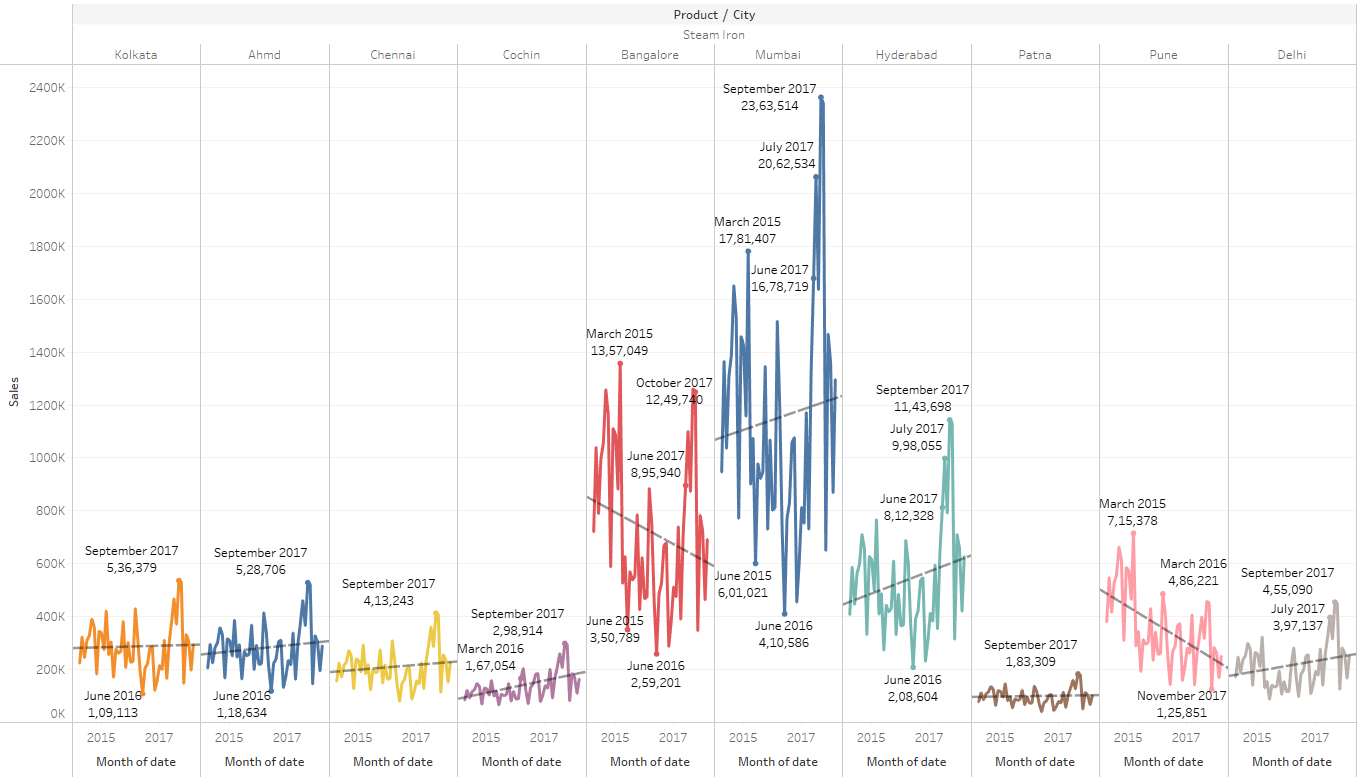
Yearly



Quarterly



Monthly



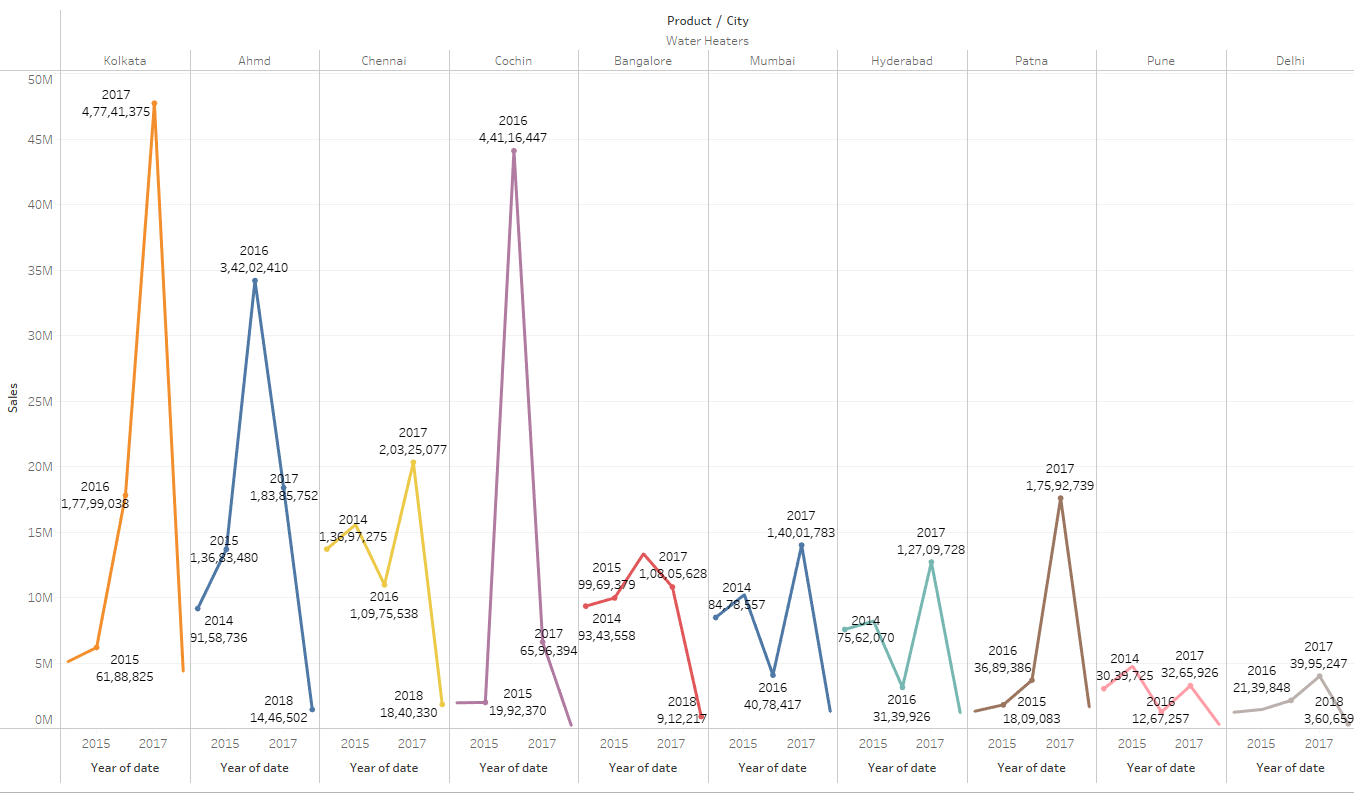
Yearly – Mumbai shows promising sales than other cities. Sales of Steam Iron is not consistent in all cities and it varies across the years

Quarterly - No pattern of increase or decrease in sales. Q3 shows highest sales across all cities

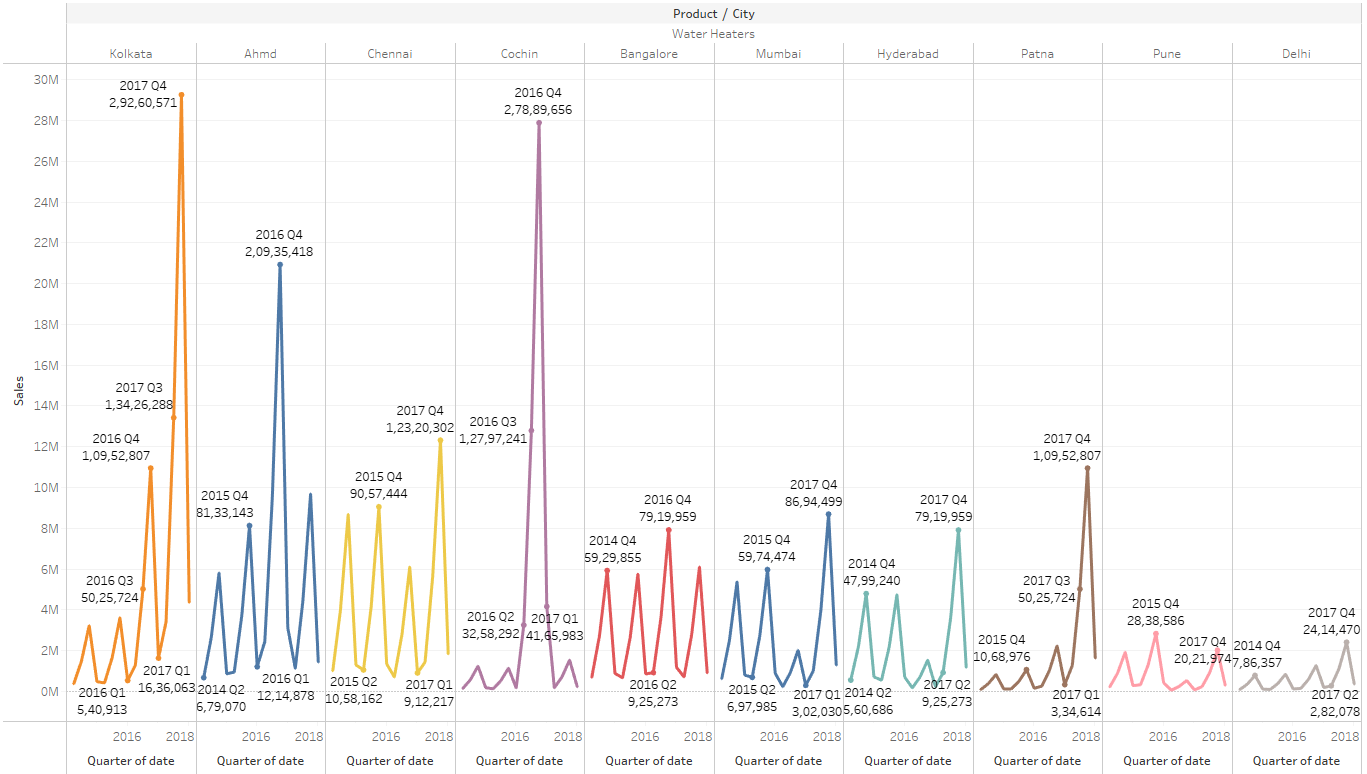
Monthly –September has the highest sales across all cities except Bangalore and pune. There is increase in trend line in Mumbai, Kolkata, Ahmedabad, Chennai ,Hyderabad , Delhi and cochin. Decrease in sales in Pune, Bangalore and flat line in Patna.

## Water Heaters vs Cities Vs Date

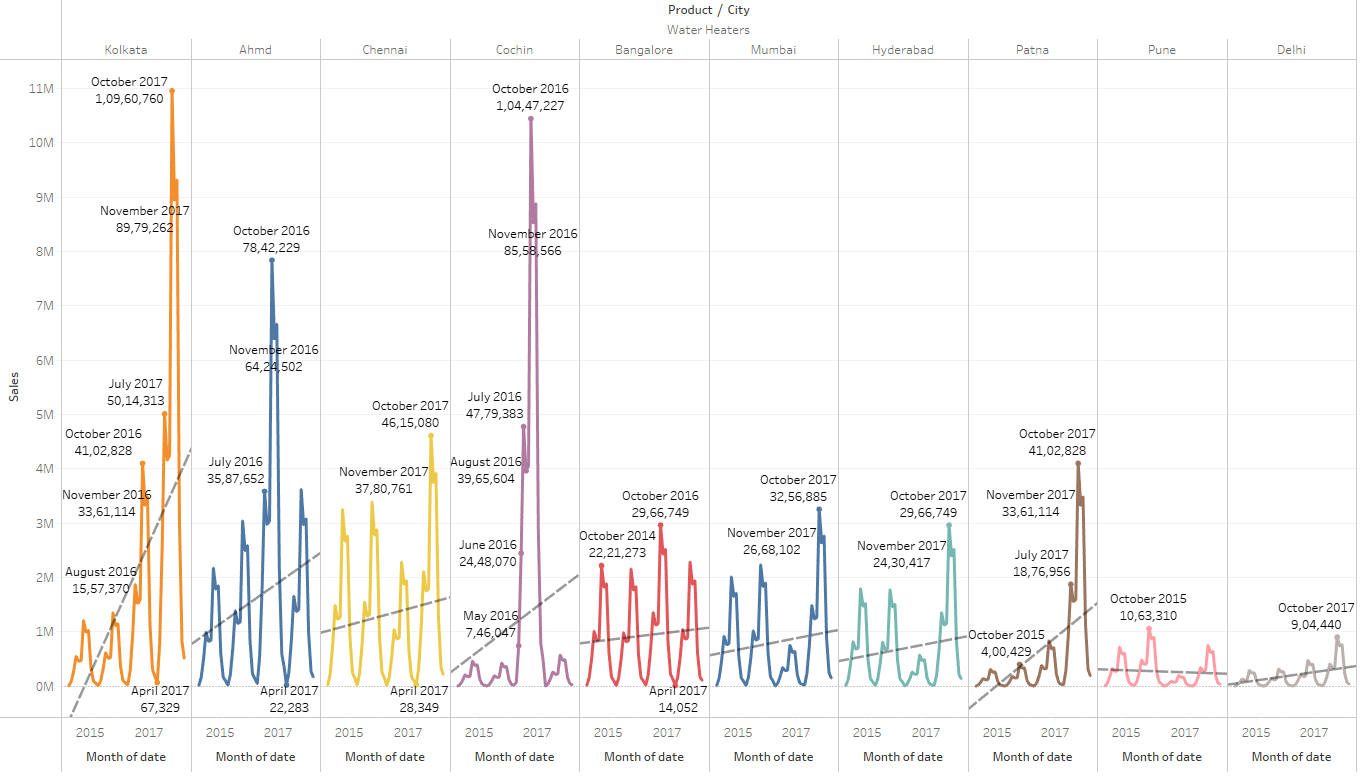
Yearly



Quarterly



Monthly



Yearly – Kolkata shows promising sales than other cities. Sales of Water heater is not consistent in all cities and it varies across years

Quarterly - No pattern of increase or decrease in sales. But sales gradually increases with slight dip in certain months except pune. Q4 shows highest sales across all cities

Monthly –October has the highest sales across all cities. There is increase in trend line in all cities except Pune.

## Missing Data and duplicates

There is no Missing data and duplicates in the dataset.

## Outlier treatment

No outlier treatment for sales since dates and sales are continuous over a period of time. Forecasting data will be affected if outliers are treated

# Business Insights

Time series is continuous so there is no Imbalance

No clustering because data has to be split based on distance and similarity. Cannot remove product and city, which are key variables in the data set

From the EDA point of view, For coolers product Chennai holds highest sales and Delhi holds lowest sales. Therefore Q1 and Q2 production should be high and rest of the quarters, production is almost zero in all cities, hence production of products can be increased and off season promotion offers can be implemented such as discounted rates or free warranty extension during off seasons sales purposes.

For Dry Iron Highest is recorded in Mumbai and lowest in Delhi. Production should be high in Mumbai and less in Delhi. Q3 records high sales, Q2 and Q4 records comparatively low sales, hence, promotions such as offering iron stand free along with the product can promote sales during low sales. Pune on the other hand declines sales, so additional promotional or discounts can be advertised to attract customers.

For Food processor Mumbai has highest sales and Patna records lowest sales. Q1 and Q3 in almost all cities records good sales. Cochin records almost flat line in sales, hence additional advertisement have to be followed in the city for awareness of the product and its utilities. In other quarters, company can give additional warranty free for promotional purpose.

For Gas Stove Again Mumbai records highest sales and Bangalore , Delhi are almost low in sales. Q3 records highest sales in almost all cities. Hence in Bangalore and Delhi , discounted prices such as 20% flat offer or exchange of old Gas stove, one can buy new gas stove for discounted rates can be offered during low sales period.

For Induction Cookers, Kolkata and Mumbai has the highest sales and Cochin and Delhi has declining sales. Awareness about induction cookers can be advertised in newspapers because this product is not widely used in all cities, hence promotional offers such as exchange of old cookers , one can get discounted Induction cookers can be offered.

For Mixers, Kolkata has highest sales and Chennai has the lowest sales. Q3 has the highest sales in all cities. Extension of motor warranty of mixers can be offered to customers or additional juicer jar can be offered free during low sales time.

For oven Toaster Grill, Mumbai records highest sales, Chennai has the lowest sales. Almost in all cities Q 4 has the highest sales. So cooking recipe books can be offered free along with the toasters and toaster gloves can also be offered along with discounted rates.

For SECF, Cochin records highest sales, and Delhi has lowest sales. Q2 records highest sales in almost all cities. Sales is good in almost all cities. Hence, additional discounted offers can be provided to the customers.

For Steam iron, Mumbai records highest sales and cochin has lowest sales. Q3 has highest sales in almost all cities and Bangalore shows a declining trend. Hence, ironing table can be promoted with discounted rate as a combo offer to the customer during low sales.

For water heaters, Kolkata and cochin has highest sales and Pune has lowest sales and there is a trend , where Q1 and Q2 has lowest sales. Hence , Warranty extension for water heaters and additional offers can be promoted.

Kolkata, Mumbai are high in population, hence high sales can be seen in these sates. Whereas, Delhi, pune and Patna sales will be low due to less population on the whole.

We will proceed to top layer of cities giving high sales in products. There are about 10 x 10 = 100 permutations and combinations possible for each product 1 x 10 cities = 10 combinations.

# **Model building**

We have explored Triple exponential Model, ARIMA Mode, SARIMA Model and Prophet Facebook Model.

ARIMA model – gave good RMSE score for Trend and Base factor

Triple Exponential Model – Gave good RMSE score for dataframe with base, trend and seasonality factor.

Model has been build based on trend and seasonality.



These are the results which have been incurred based on model building for the top 4 cities in top selling 8 products. SECF and Oven Toaster Grill are the least selling products.

# Forecasting from the modelling results

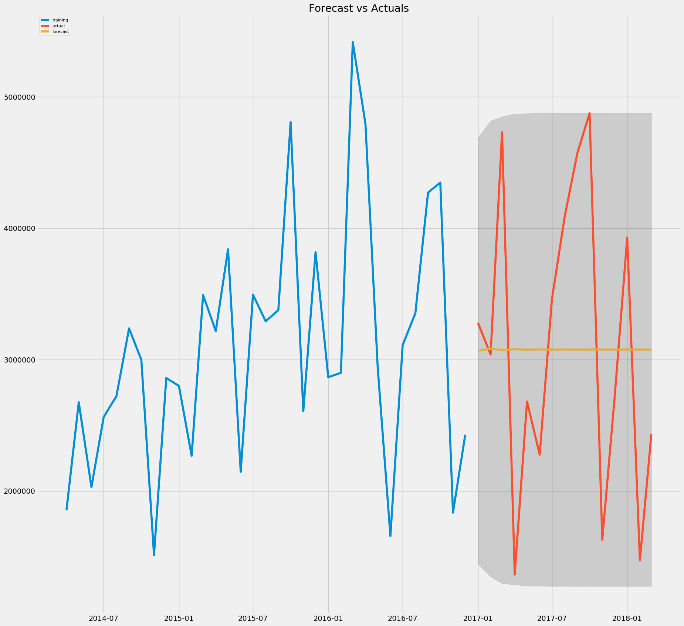
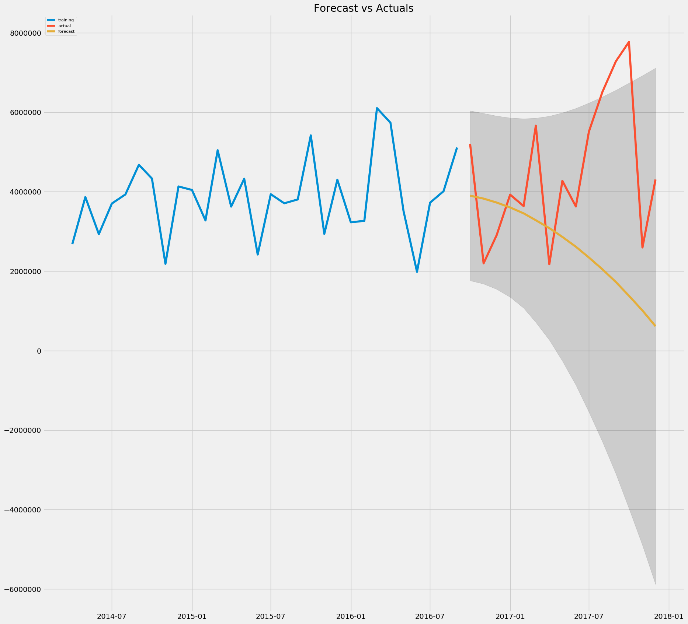
## Mixers – Mumbai

Constant Forecast



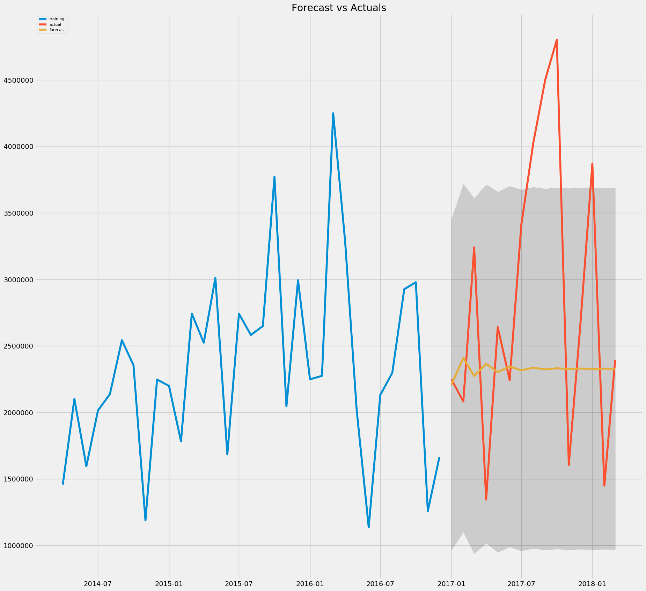
## Mixers – Kolkata Mixers – Bangalore

Downward Forecast Constant Forecast

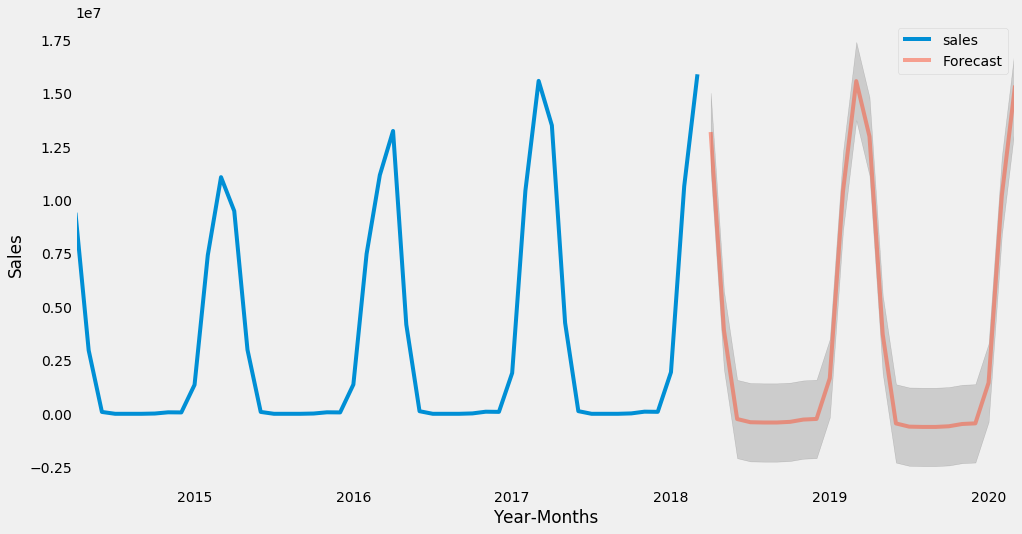


## Mixers – Hyderabad

Constant Forecast.

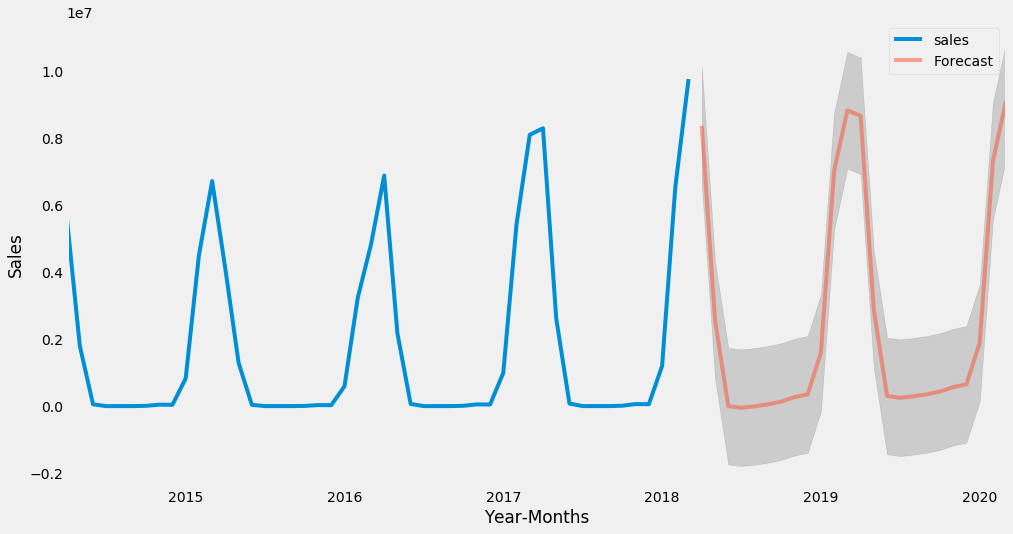


## Coolers – Chennai

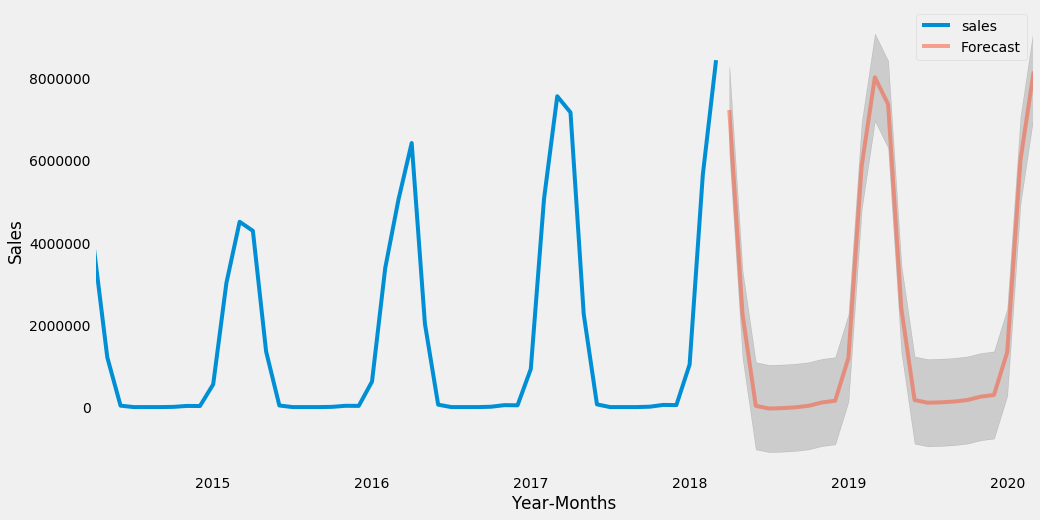
Seasonal Forecast with increasing Forecast. 

## Coolers- Mumbai

Seasonal Forecast with Constant Forecast.

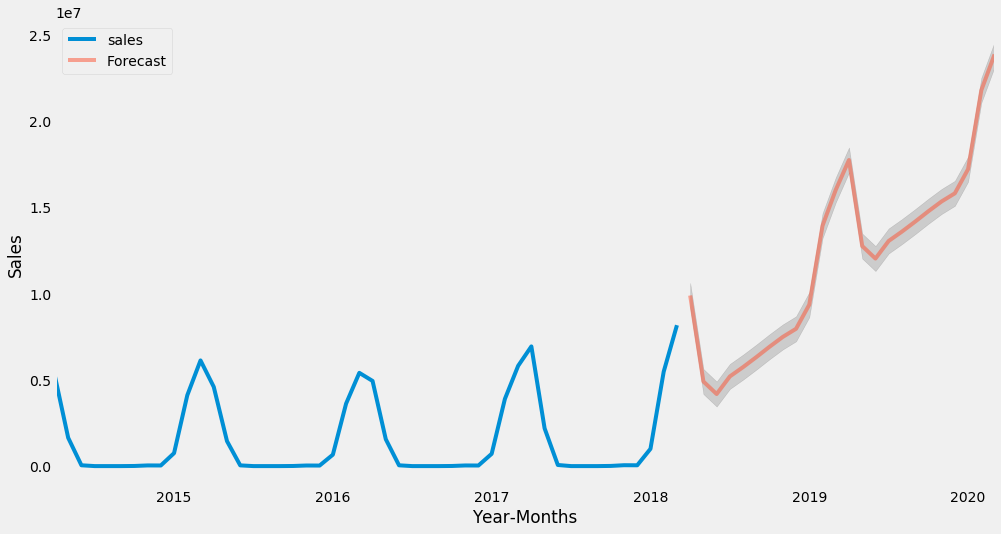


## Coolers – Cochin

Seasonal Forecast with Constant Forecast. 

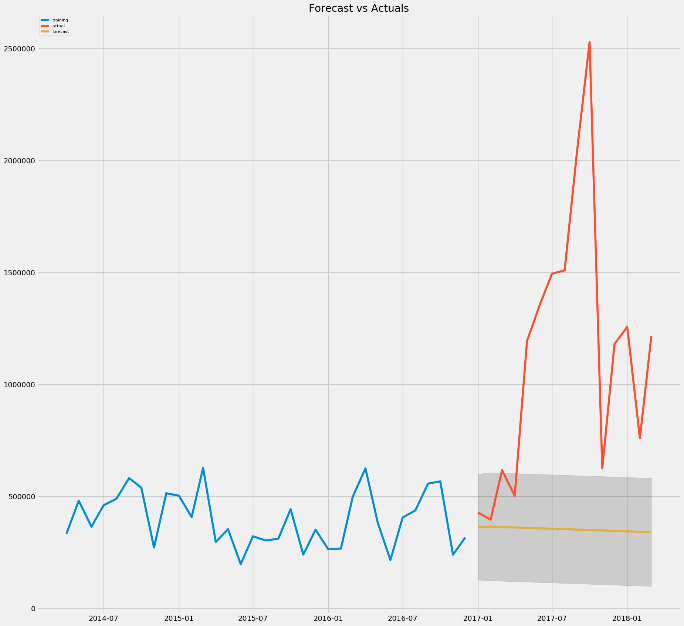
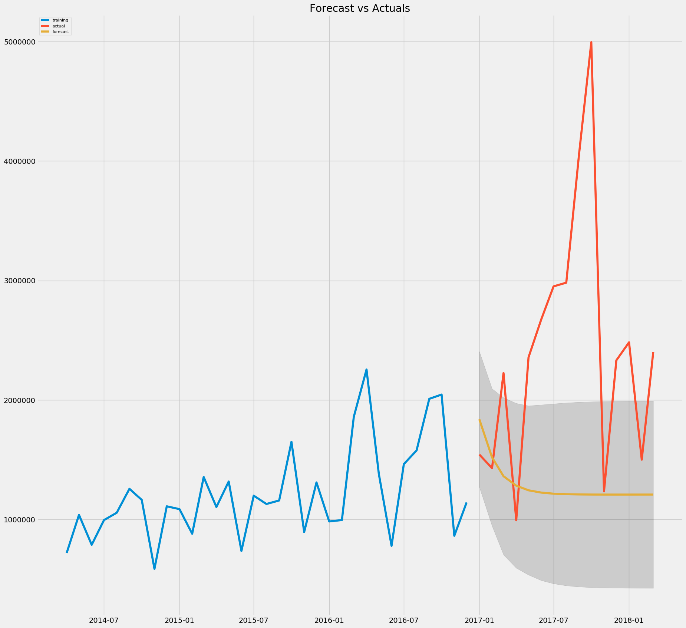
## Coolers – Bangalore

## Seasonal Forecast with Constant Forecast.



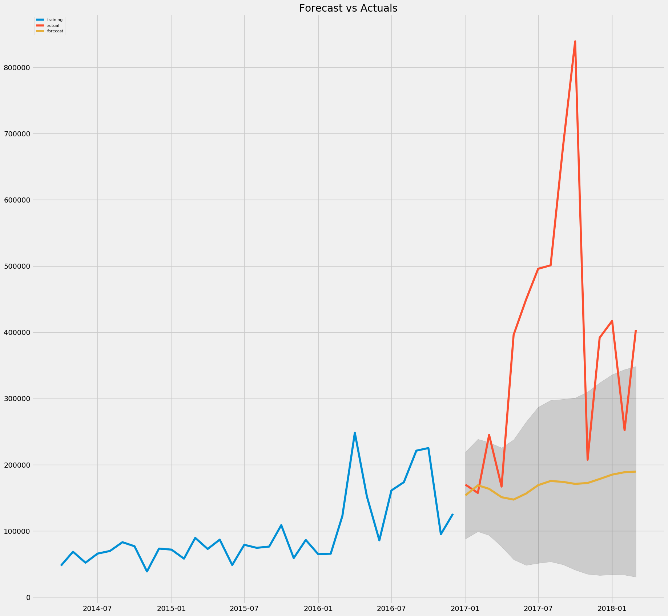
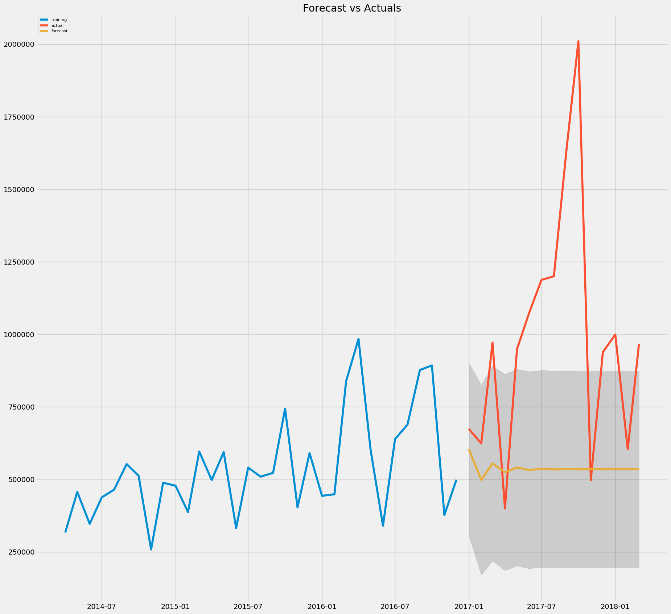
## Dry Iron – Mumbai Dry Iron – Kolkata

Constant Forecast Constant Forecast



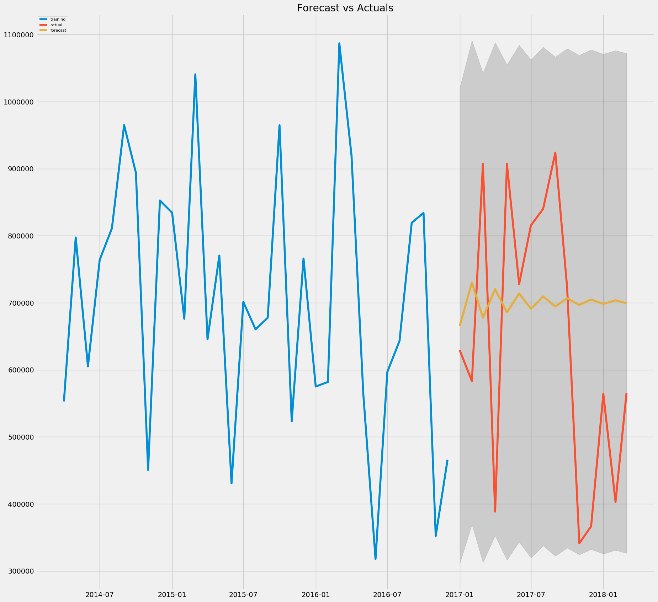
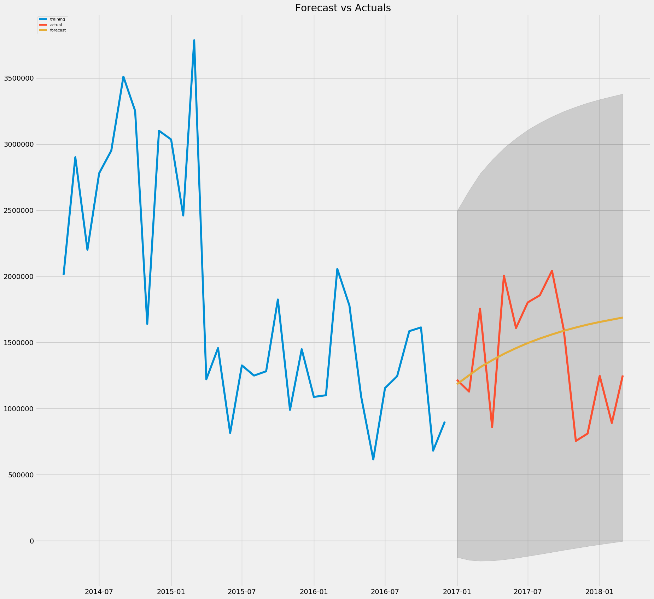
## Dry Iron – Hyderabad Dry Iron – Patna

Constant Forecast Increasing Forecast



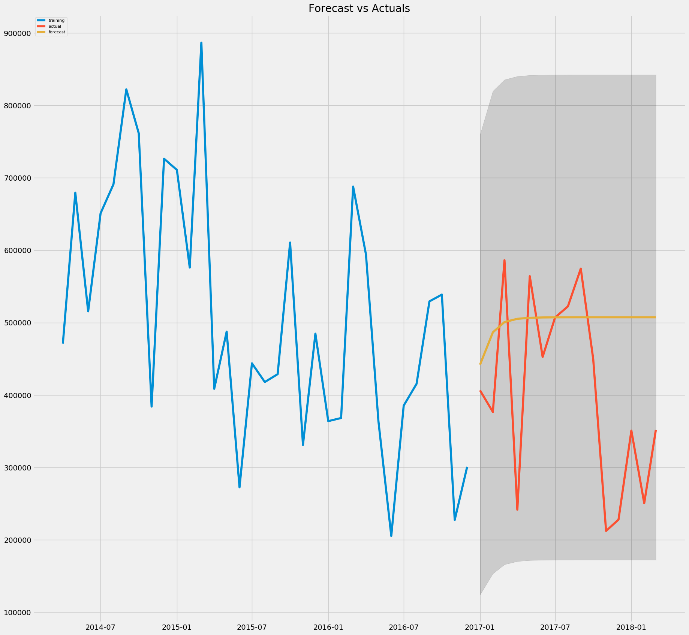
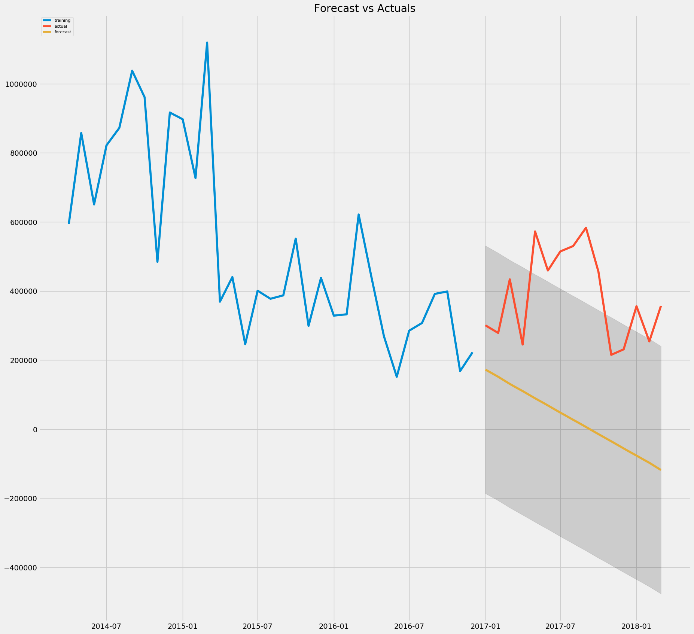
Food Processor – Mumbai Food Processor – Pune

Increase Forecast Constant Forecast



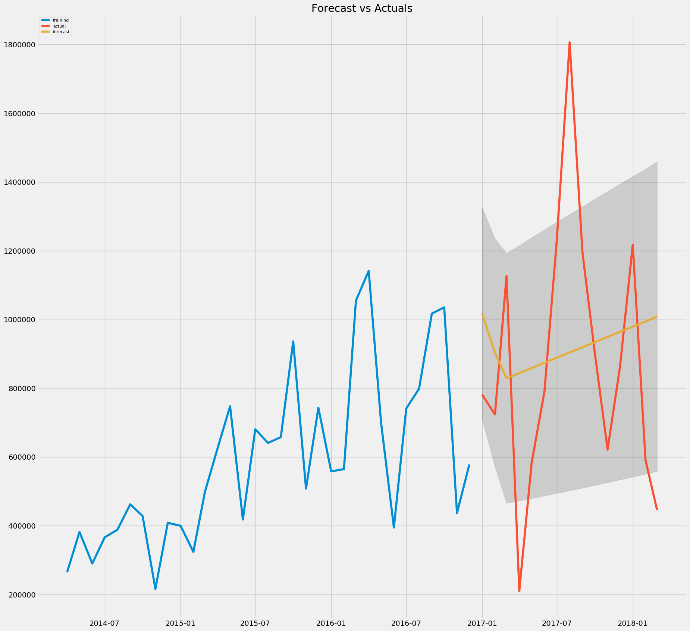
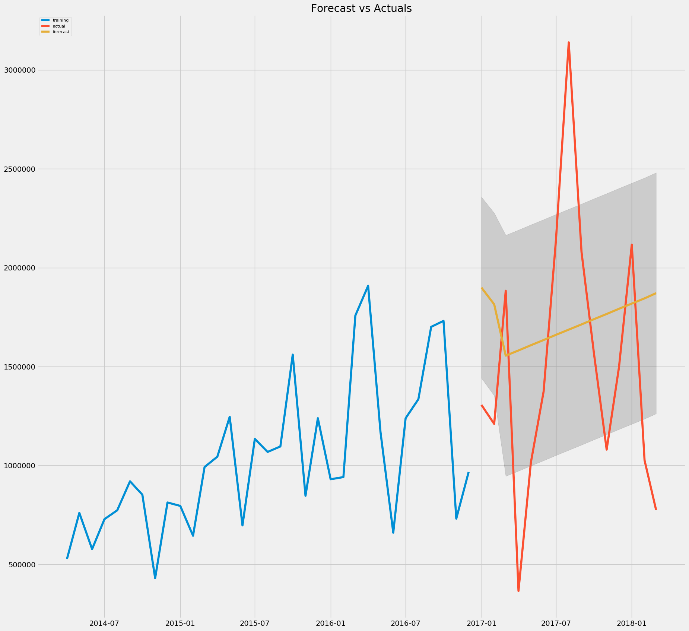
Food Processor – Bangalore Food Processor – Ahmd

Downward Forecast Constant Forecast



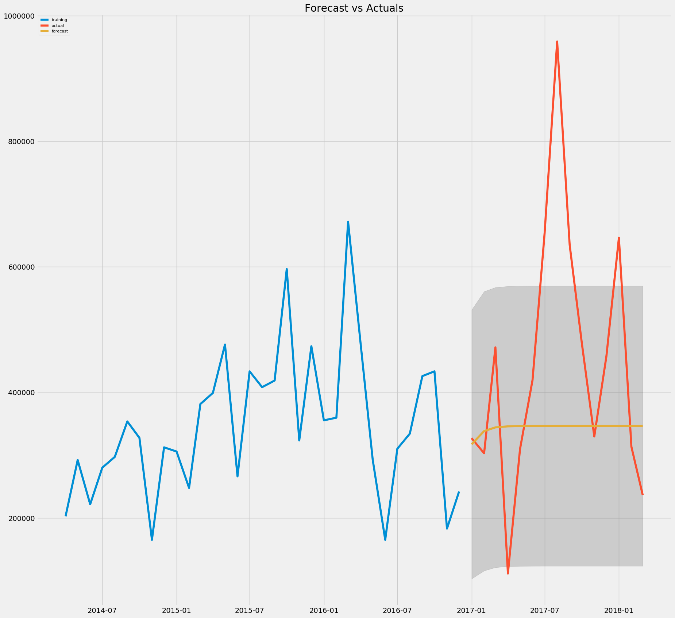
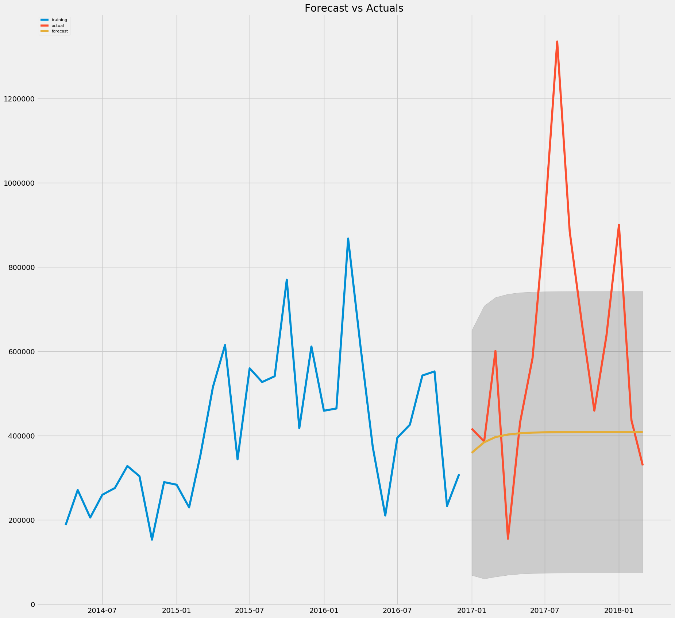
Gas Stove – Mumbai Gas-Stove - Cochin

Increasing Forecast Increasing Forecast



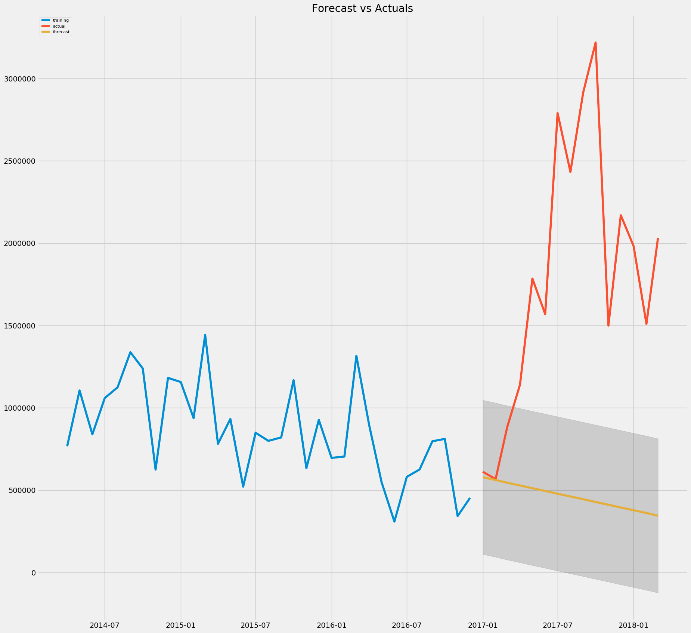
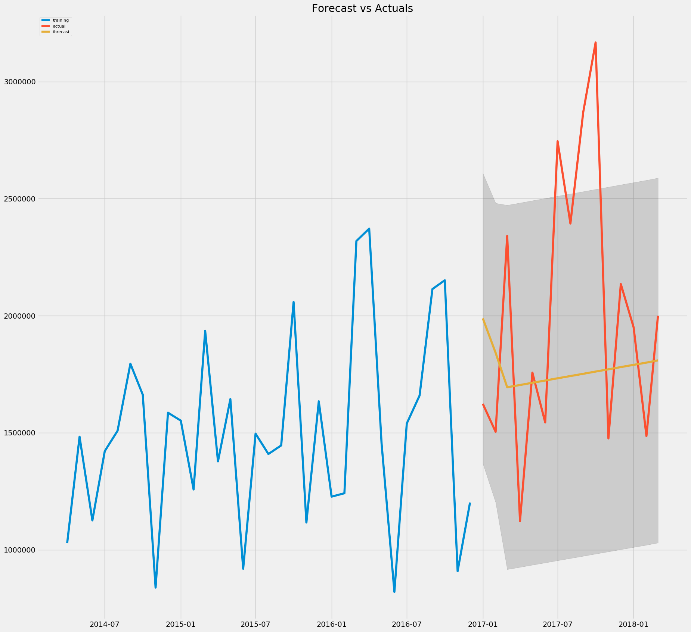
Gastove- Kolkata Gastove- Hyderabad

Constant Forecast Constant Forecast



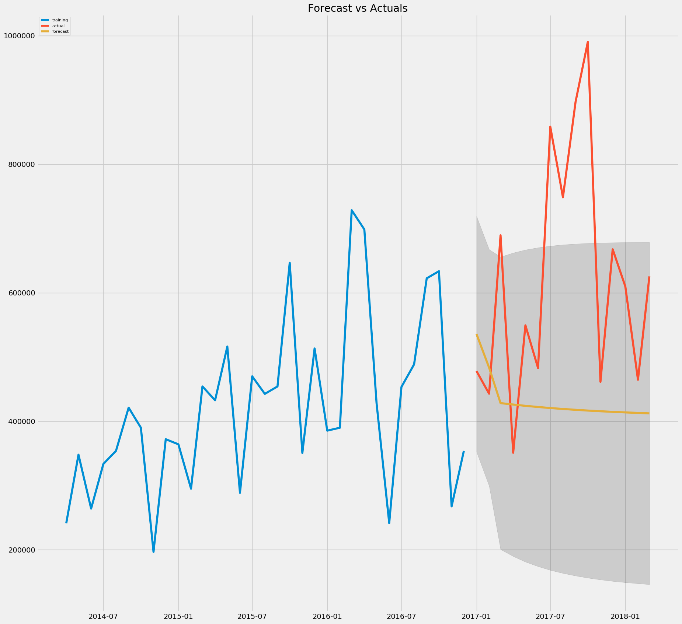
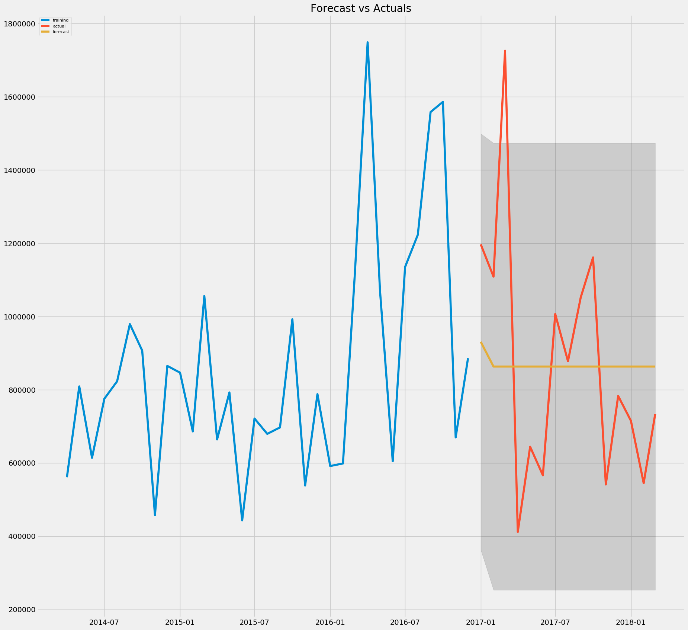
Induction Cookers – Mumbai Induction Cookers – Mumbai

Increasing Forecast Downward Forecast



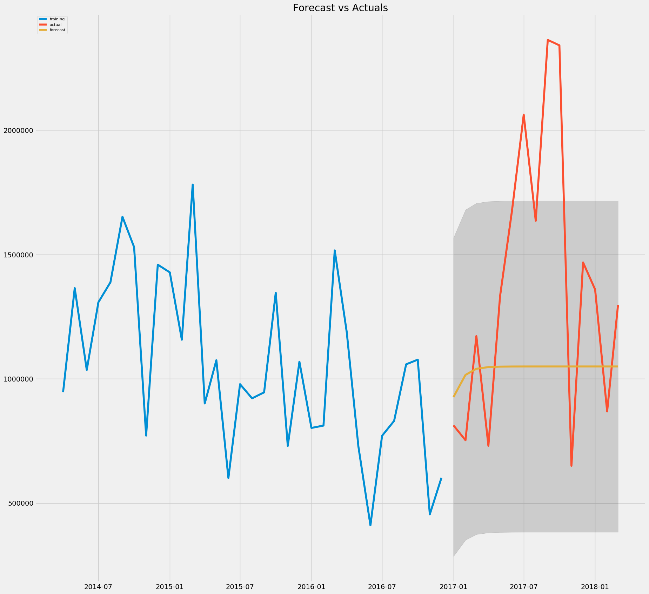
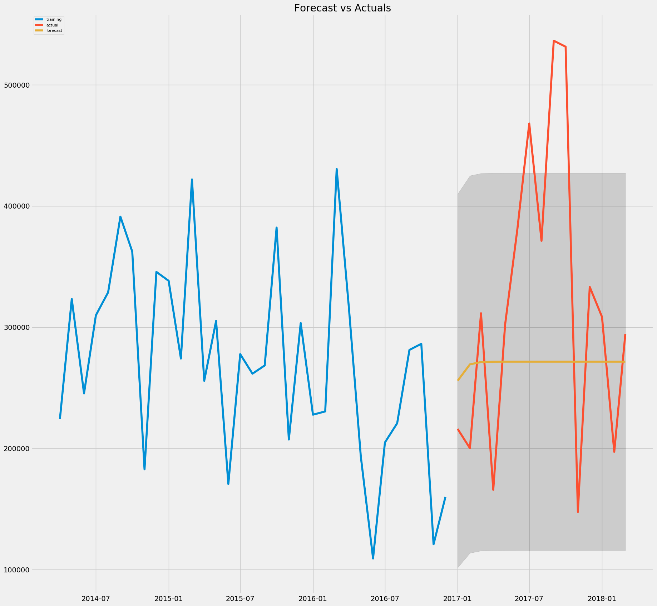
Induction Cookers – Bangalore Induction Cookers – Hyderabad

Constant Forecast Constant Forecast



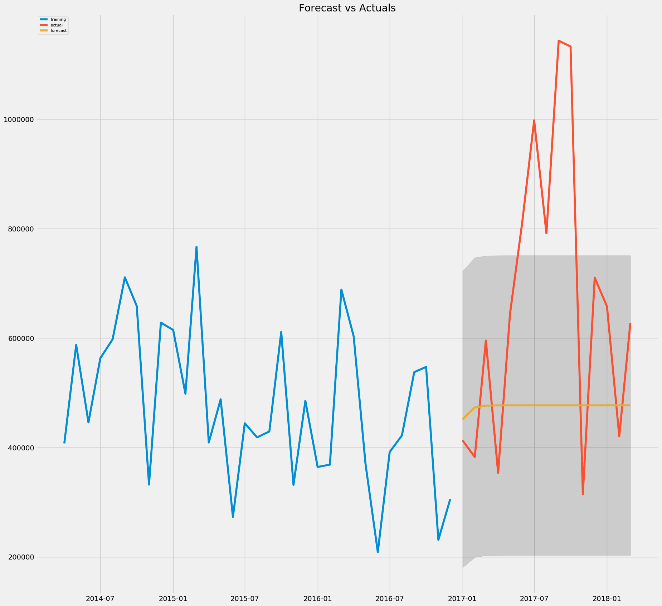
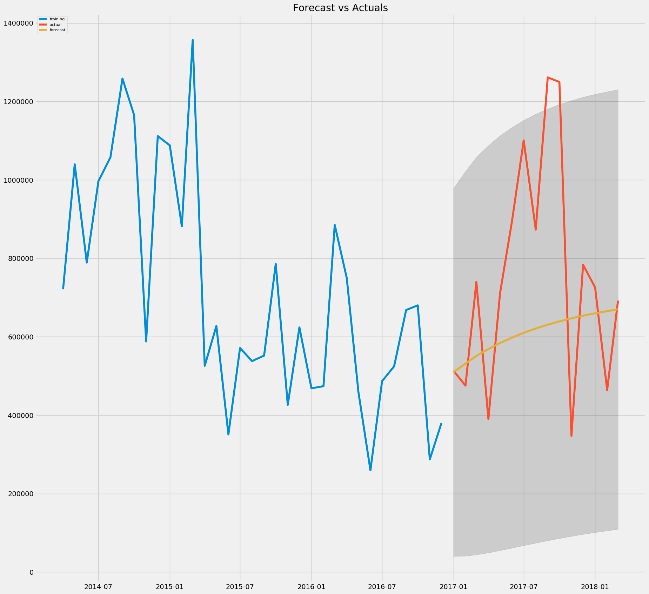
Steam Iron – Mumbai Steam Iron - Kolkata

Constant Forecasting Constant Forecasting

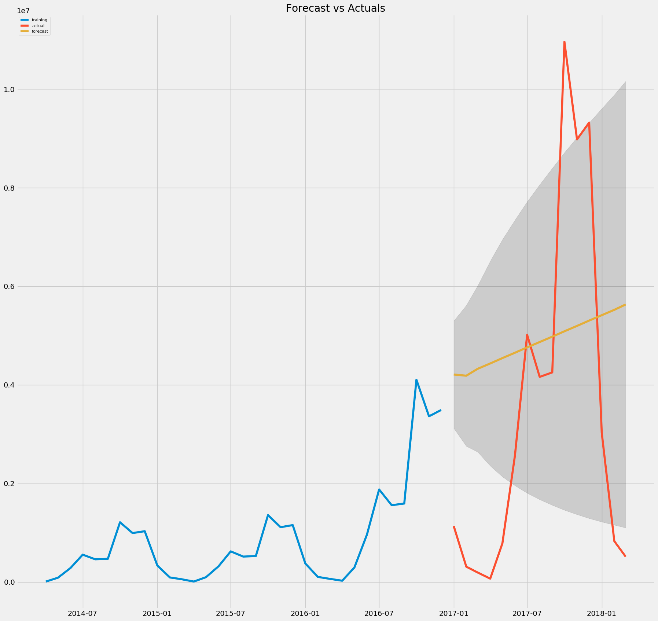
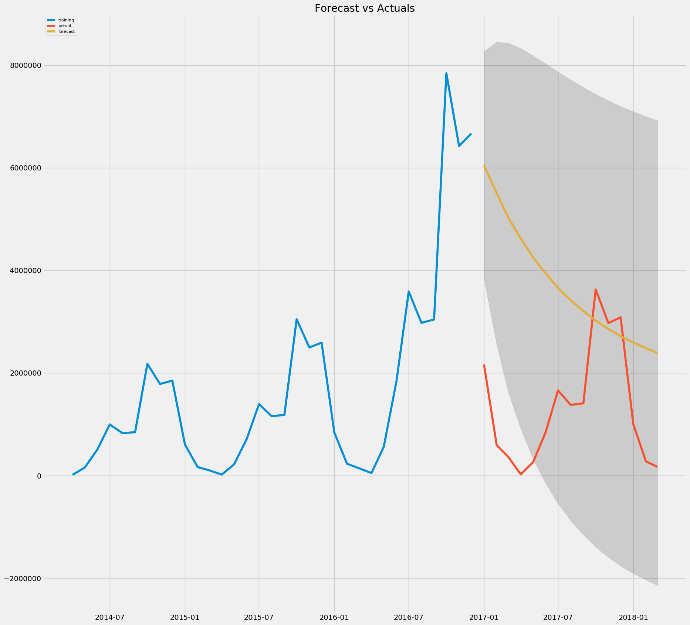
Steam Iron – Bangalore Steam Iron - Hyderabad

Increasing Forecast Constant Forecast



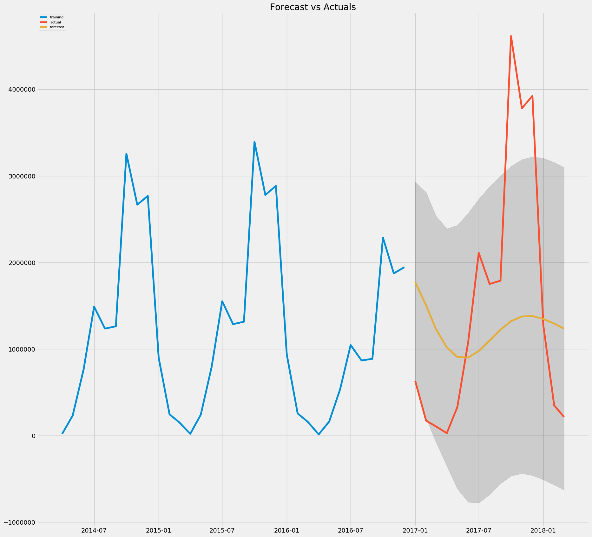
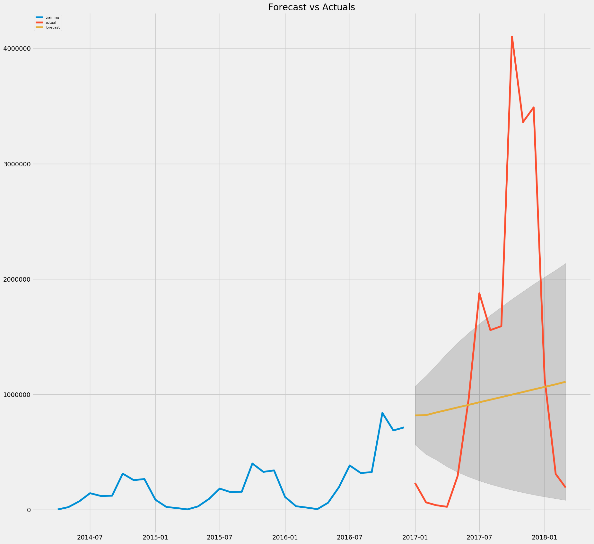
Water Heater – Ahmd Water Heater - Kolkata

Downward Forecast Upward Forecast



Water Heater – Patna Water Heater - Chennai

Upward Forecast Downward Forecasting



Model accuracy we have explored facebook Prophet, SARIMA and ARIMA models for accuracy.

We could also explore Hierarchical and Groped Time series model.

# Recommendations and Insights

* Based on the above results , one can divide the forecasting into Upward, constant and downward forecasting.
* Constant – Explore Growth Opportunities
* Upward – Sustain and keep the growth upward
* Decreasing – Needs attention.
* Delhi although produces least sales, there are upward trend in some cities, hence logistic, sales and production strategies should be altered altogether to produce stable sales.
* SECF also have an upward sales trend overall, hence logistic, sales and production strategies should be altered altogether to produce stable sales.

## Explore Growth Opportunities

* Explore new regions for business
* Enhance Sales channels, improve dealership and provide distribution offers
* New billing systems such as self billing.
* Explore online platforms such as applications and advertise on social media
* Explore competitor’s pricing and discounts
* Equip new features according to the customer’s demand
* Check the product in other regions providing same sales
* Seasonal offers, discounts to grab the attention of the customer
* Provide additional training for staffs about sales
* Additional hire part time staff during peak hours
* Avoid delays due to system failure.

## Sustain and keep the growth upward

* Maintain high volume of stock and inventory across sales channel
* Periodic check on existing dealers and distributors
* Maintain current pricing strategy and provide innovative offers
* Maintain or hire staff without attrition
* Calculate the number of customers monthly and maintain healthy customer ratio
* Maintain same logistics strategies and improvise if necessary
* Maintain high profit margin to maintain sales
* New warehouses can be build for easy access to high selling stores

## Need Attention

* Audit store level and deep drive in to reasons for decreasing sales
* Change distributors and dealers if required
* Check product trend and features with competitors in the market
* Check the quality products
* Revise marketing strategies , sales and logistics
* Combine offers along with high selling products with low selling products
* Understand the behaviour of the market by consumer buying patterns
* Open online stores and free delivery to customers
* Seasonal offers and holiday offers to attract customers
* Extended period of warranty with minimal amount to customers