BUDGET SALES ANALYSIS

DETAILED PROJECT REPORT

PRANIT PATIL

iNeuron

PROJECT DETAIL

Project Title	Budget Sales Analysis
Technology	Business Intelligence
Domain	Sales and Retail
Programming Language Used	Python
Tools Used	Jupyter Notebook, MS-Excel, MS-Power BI

OBJECTIVE

This project is to analyse the Customer data, Product data, Sales data and Budget Data of a Retail chain and extract keys insights that can be valuable in taking business decisions.

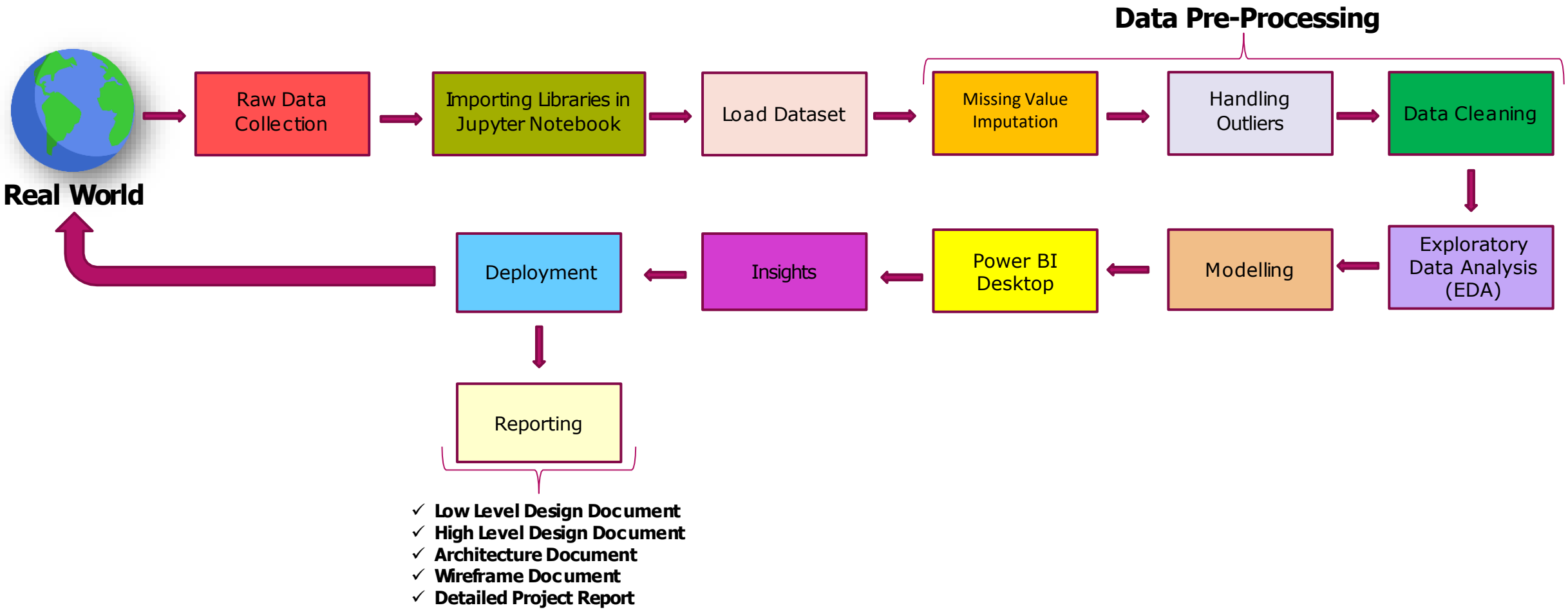
❑ Benefits:

- Help in making wiser business decisions.
- Aid in customer satisfaction and trend monitoring, which can serve current consumers and attract new ones.
- Greater client base understanding is provided.

PROBLEM STATEMENT

Budget and Sales are by far most important attributes that defines a business's success and failure. Therefore, it is very important to keep a track on various features related to these attributes to keep on increasing the Sales and to allocate the Budget so that it can be utilized wisely and efficiently. So, it is very important for businesses to dig deep into the customer, sales, budget and product data to make better marketing strategy, to know the target customers, to make market friendly product upgrades and to keep a strong track on the budget efficiency. Good data driven systems can help achieve these goals and take the businesses forward towards success.

ARCHITECTURE



DATASET INFORMATION

Customer Data:
This file consists of the features related to the data about the customers i.e. ‘CustomerKey’, ‘FirstName’, ‘LastName’, ‘FullName’, ‘BirthDate’, ‘MaritalStatus’, ‘Gender’, ‘YearlyIncome’, ‘TotalChildren’, ‘NumberChildrenAtHome’, ‘Education’, ‘Occupation’, HouseOwnerFlag’, ‘NumberCarsOwned’, AdressLine1’, ‘DateFirstPurchase’, ‘CommuteDistance’.

Product Data:
This file consists of the features related to the data about the product i.e. ‘ProductKey’, ‘ProductName’, ‘Subcategory’, ‘Category’, ‘StandardCost’, ‘Color’, ‘List Price’, ‘DaysToManufacture’, ‘ProductLine’, ‘ModelName’, ‘Photo’, ‘ProductDescription’, ‘StartDate’.

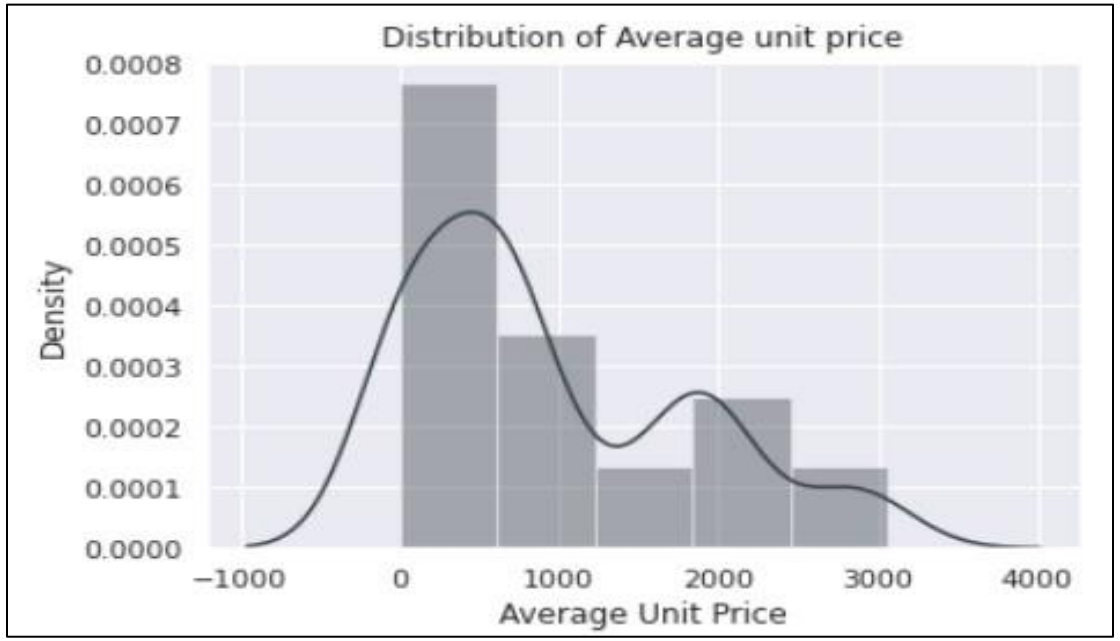
Sales Data:
This file consists of the features related to the data about the Sales i.e. 'ProductKey', 'OrderDate', 'ShipDate', 'CustomerKey', 'PromotionKey', 'SalesTerritoryKey', 'SalesOrderNumber', 'SalesOrderLineNumber', 'OrderQuantity', 'UnitPrice', ’TotalProductCost', 'SalesAmount', 'TaxAmt.

Territory Data:
This file consists of the features related to the data about the Territory i.e. ‘SalesTerritoryKey’, ‘Region’, ‘Country’, ‘Group’, ‘RegionImage’.

Budget Data:
This file consists of the features related to the data about the Budget 2016 i.e., 'Category', 'Subcategory', 'ProductName', 'ProductKey', 'Jan, 2016', 'Feb, 2016', 'Mar, 2016', 'Apr, 2016', 'May, 2016', 'Jun, 2016', 'Jul, 2016', 'Aug, 2016', 'Sep, 2016', 'Oct, 2016', 'Nov, 2016', 'Dec, 2016', 'Grand Total’.

INSIGHTS

Product Price per unit Distribution

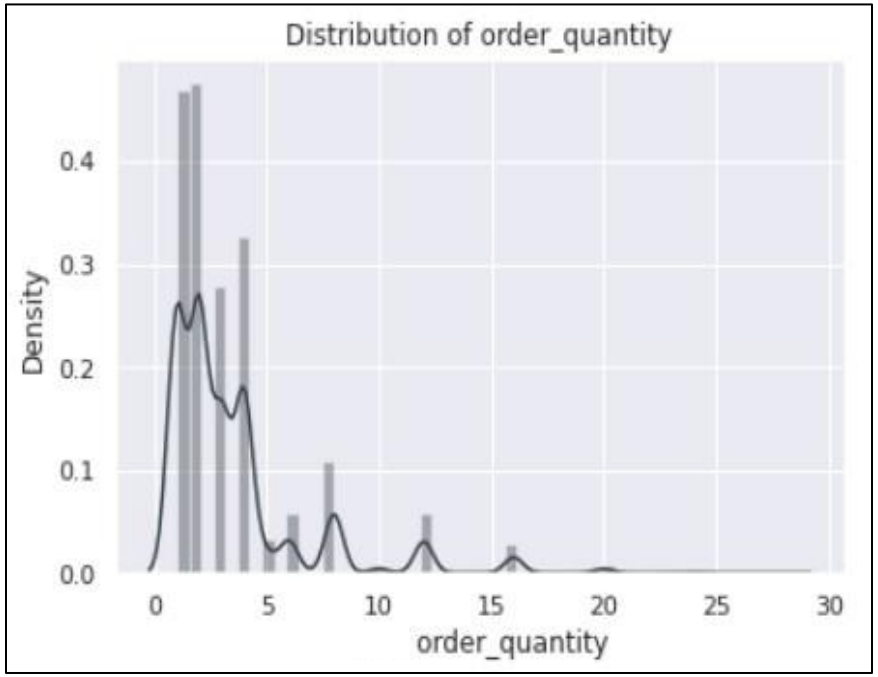


Sales order line number distribution

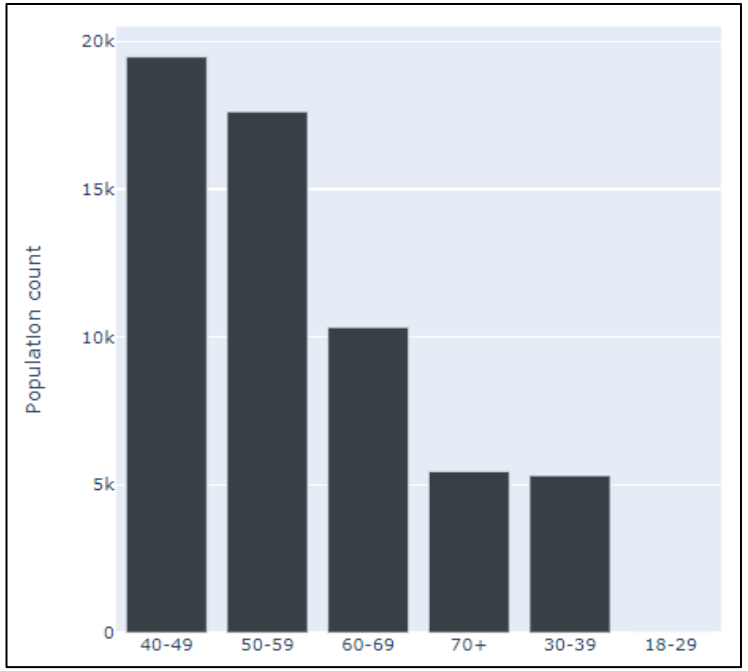
- According to the above distribution plot, we can conclude that maximum of the product unit price is below $1000

- Most of the time three to two products are ordered in a single order.

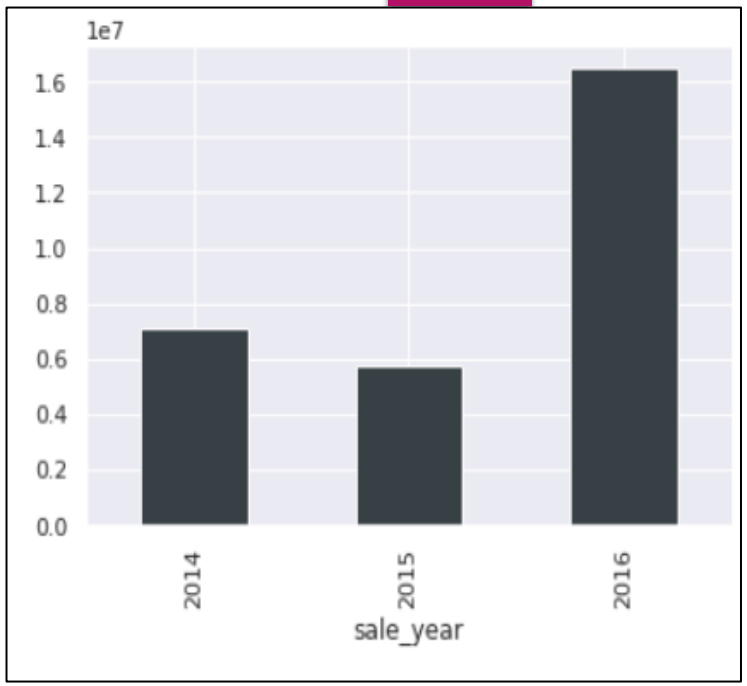
Sales order quantity distribution



Age Distribution



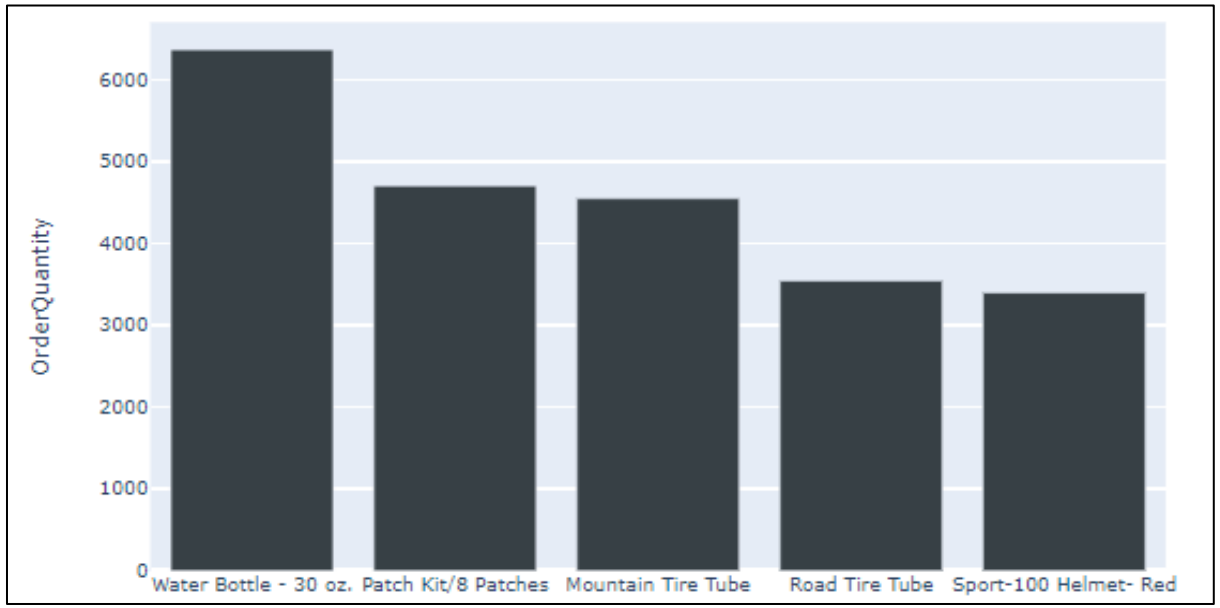
Year wise Sales



- Maximum quantity ordered for a product is below 5.
- A sizable portion of the clientele is made up of people between the ages of 40 and 59.
- The year 2016 saw an exponential surge in sales.

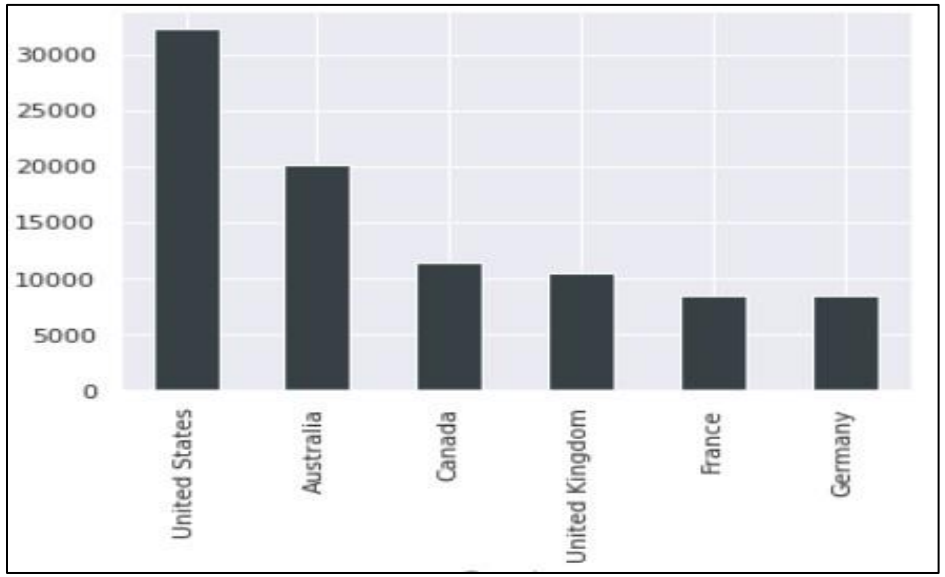
Quantity ordered based on category and subcategory from 2014 to 2016

sale_year	Category	SubCategory	OrderQuantity
2014	Bikes	Mountain Bikes	616
		Road Bikes	2876
2015	Bikes	Mountain Bikes	1661
		Road Bikes	3284
		Bike Racks	493
		Bike Stands	394
	Accessories	Bottles and Cages	12055
		Cleaners	1381
		Fenders	3239
		Helmets	9685
		Hydration Packs	1124
		Tires and Tubes	25518
2016	Bikes	Mountain Bikes	5490
		Road Bikes	6535
		Touring Bikes	3410
	Clothing	Caps	3178
		Gloves	2143
		Jerseys	5068
		Shorts	1491
		Socks	856
		Vests	824



Top 5 selling products

Country wise quantity ordered



High quantity of products is ordered from Australia and United States.

Major Profit is contributed by the Bike Category.

Overall profit based on order year, category and subcategory

sale_year	Category	SubCategory	profit
2014	Bikes	Mountain Bikes	586874.557600
		Road Bikes	2256280.998300
2015	Bikes	Mountain Bikes	1019388.334900
		Road Bikes	1375064.915000
2016	Accessories	Bike Racks	23136.960000
		Bike Stands	23689.092000
		Bottles and Cages	34448.978300
		Cleaners	4299.868800
		Fenders	27711.633000
		Helmets	135167.732700
		Hydration Packs	24303.132200
		Tires and Tubes	144793.083200
	Bikes	Mountain Bikes	2907361.198000
		Road Bikes	1905953.736400
		Touring Bikes	1454872.695900
	Clothing	Caps	4331.831500
		Gloves	20895.744100
		Jerseys	37965.228300
		Shorts	41973.524600
		Socks	3055.841100
		Vests	20948.777000

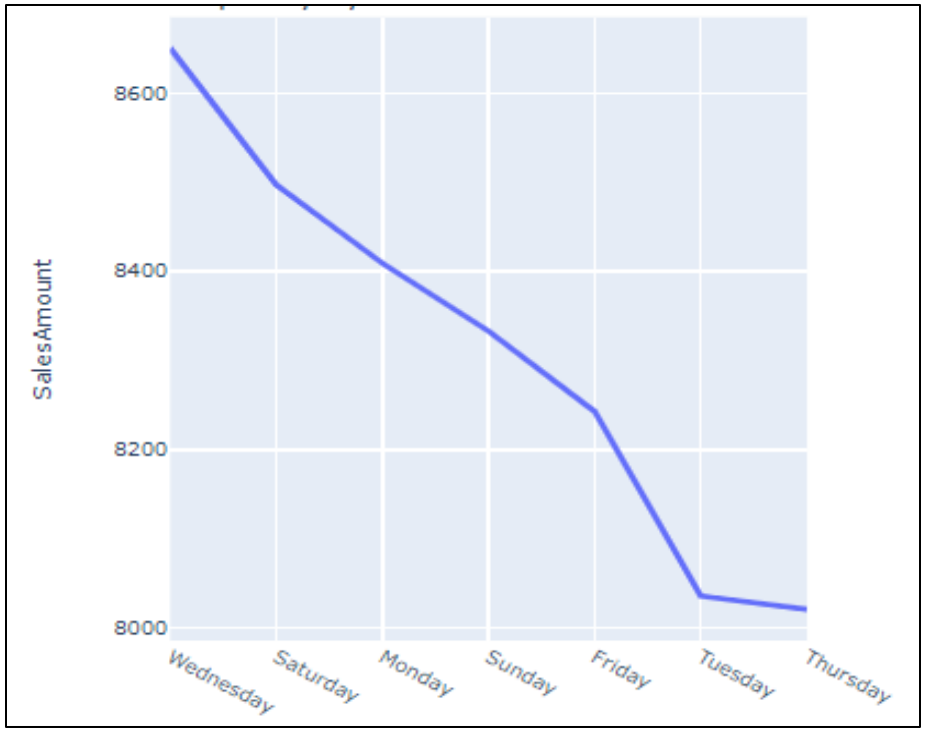
What was the best month for sales? How much was earned that month?

- Maximum profit earned in the months of **June, November, and December.**

How efficient are the logistics?

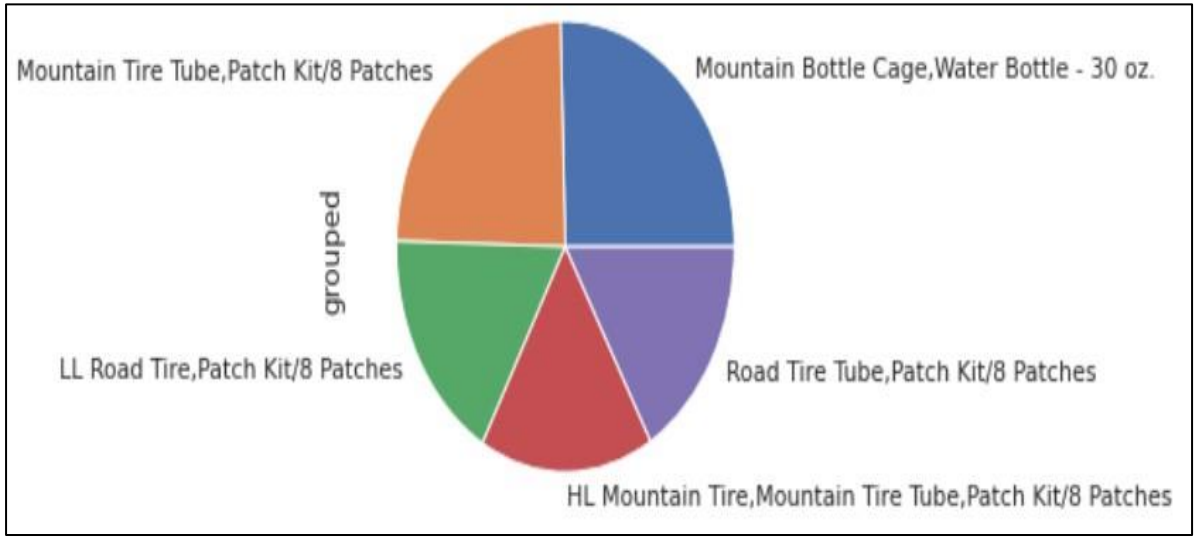
- The average order has a gap of 7 days between the day the order is ready for export from the factory and the date was shipped.
- Management must work to reduce this gap toward 3 days.

What time should we display advertisement to maximize likelihood of customer is buying product?



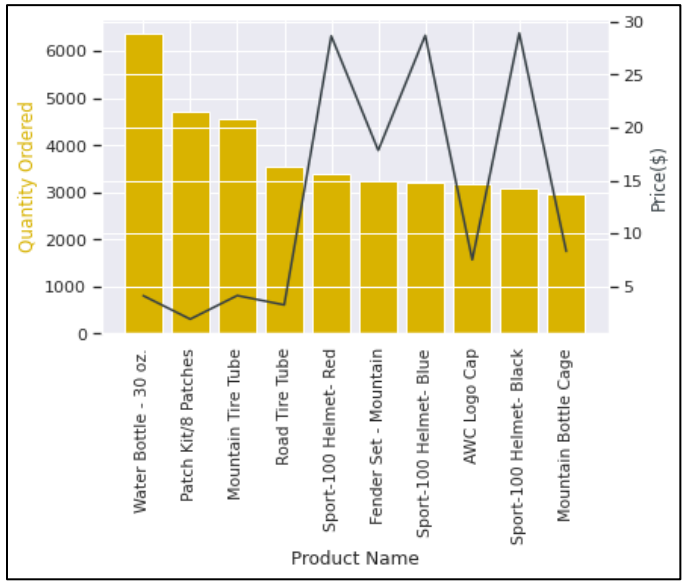
- High sales orders are seen on **Wednesday and Saturday**; therefore, we can promote our product during this workweek.

Which products are most often sold together?



- The above product can be sold in a bundle or a combined package for a discount.

Which product sold the most? why do you think it sold the most?

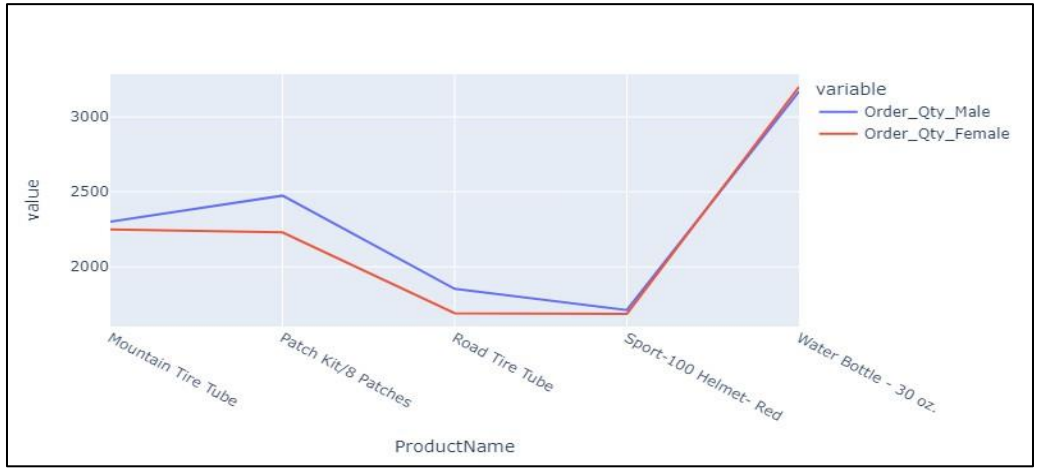



```
prices.corr(quantity_ordered)
```

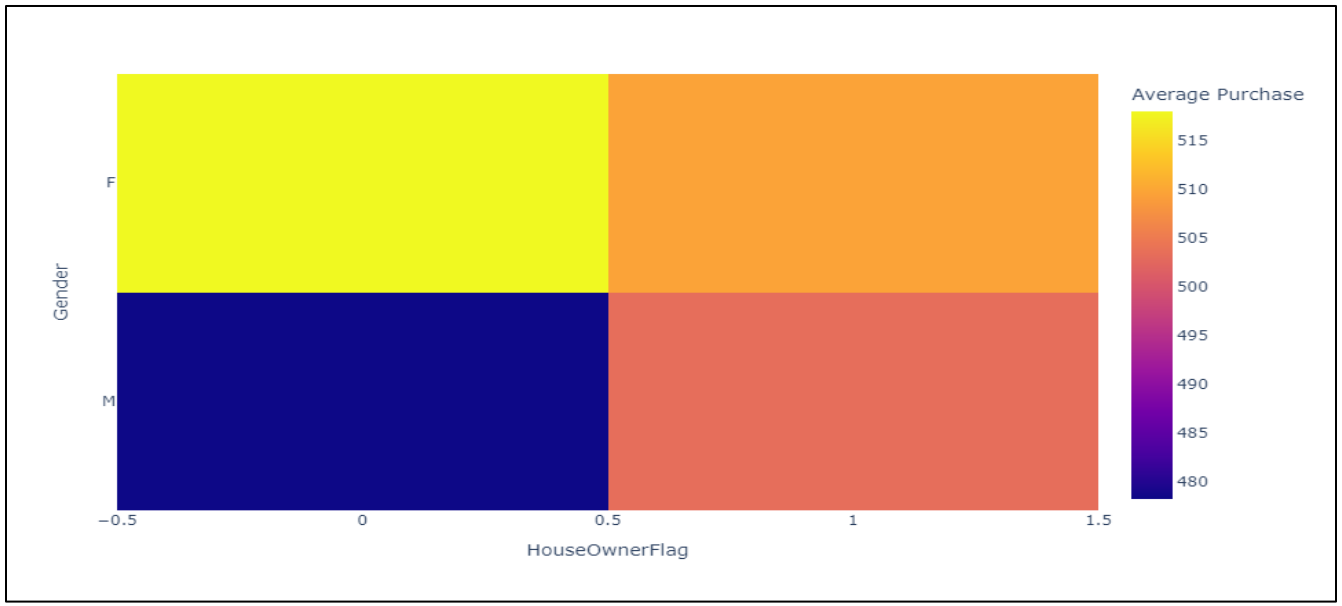
```
-0.5333019792658484
```

- There is a high negative correlation between Price and the number of Quantity ordered.
- We can conclude that low price product has a high demand.

Compare most ordered product by gender

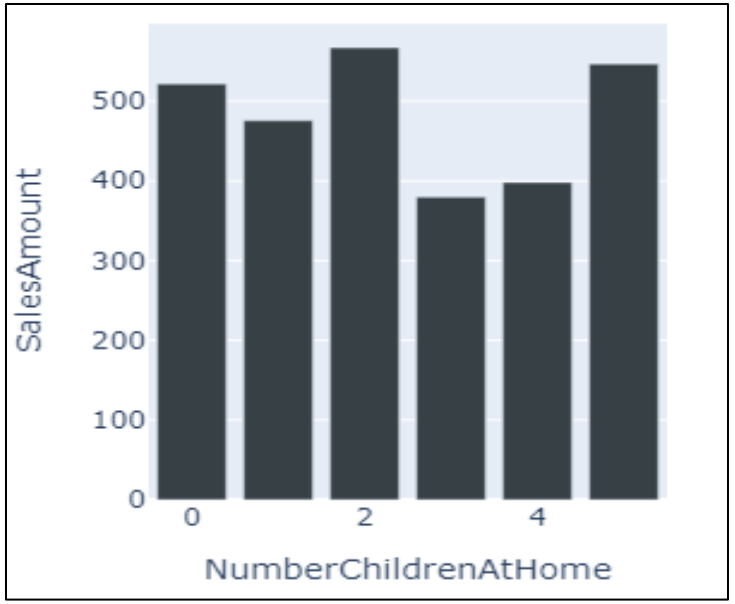


Does Gender and home ownership matter in order purchasing



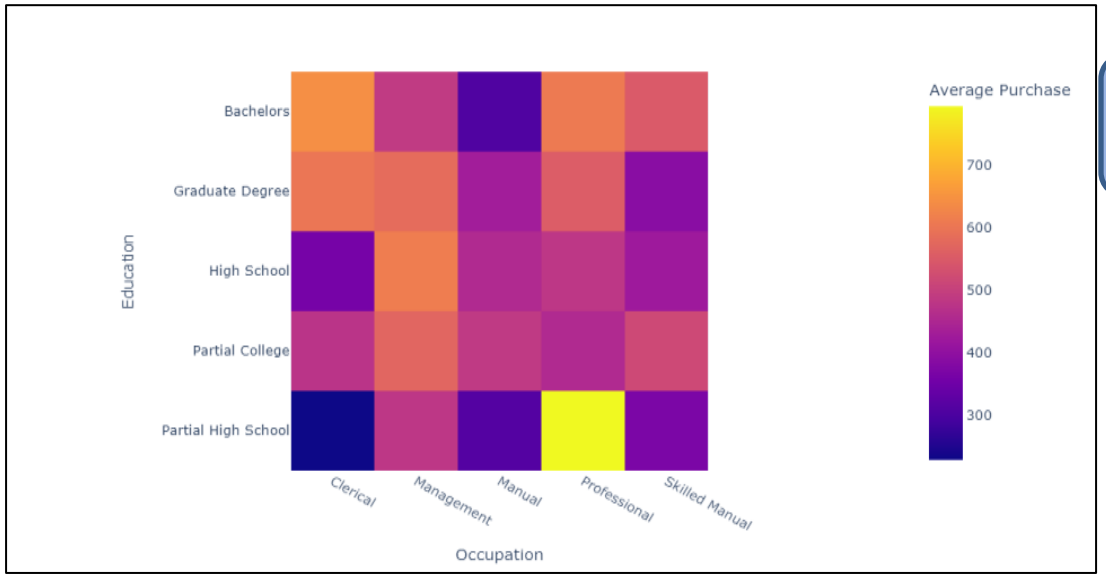
- It's interesting to note that the average amount spent by men without permanent addresses is low, whilst the average amount spent by women without permanent addresses is higher.

Number of children and Purchase correlation



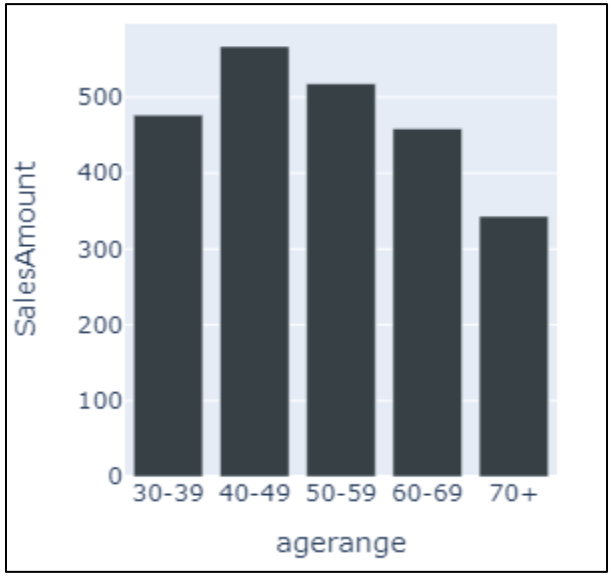
- Purchase among customers with a number of children 2 and 5 is high.

Occupation and purchase correlation



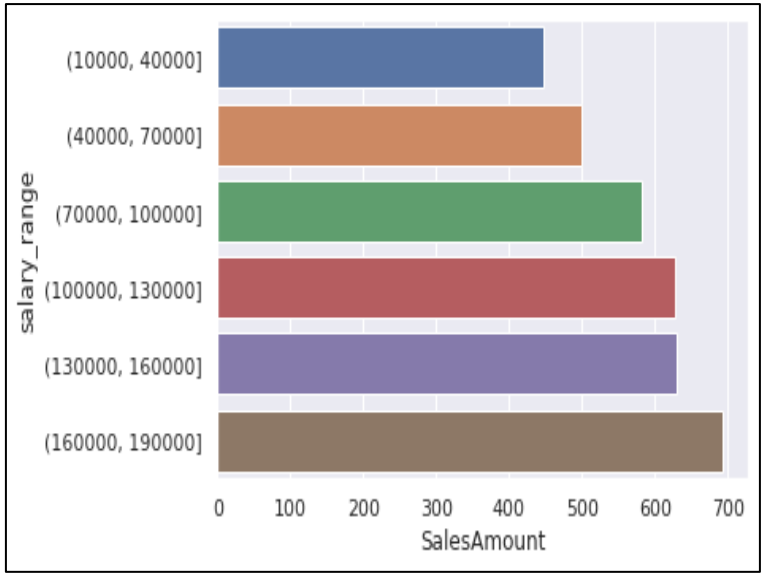
- Purchases by Professional and Management customers are comparatively high.

Which age group has produced the most revenue?



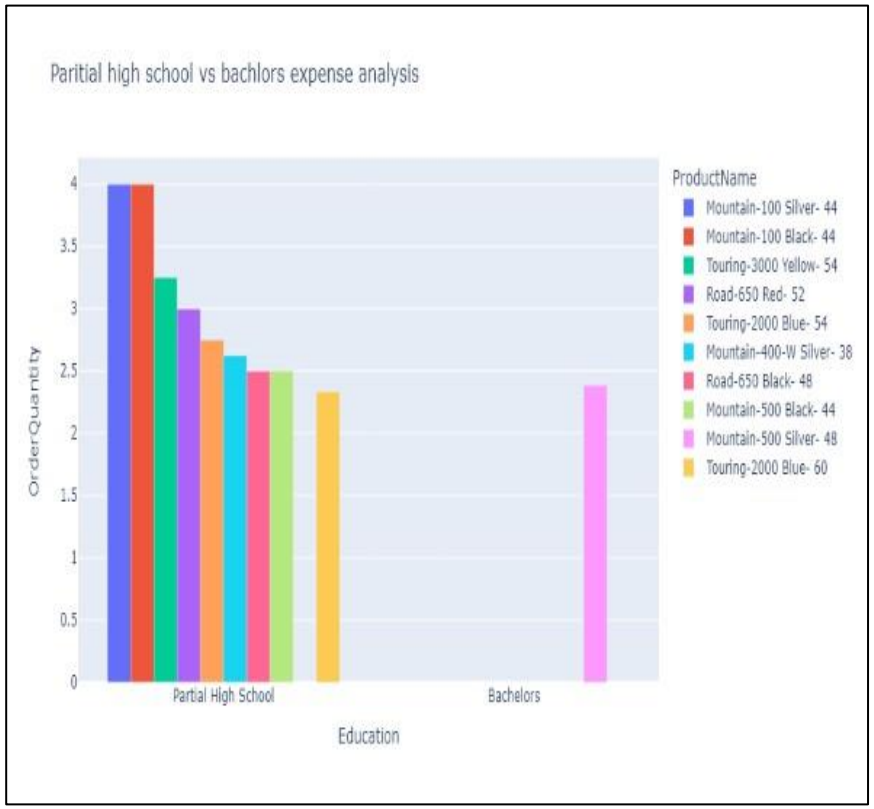
- Age range of 40-49 and 50-59 shows high demand compared to another age group.

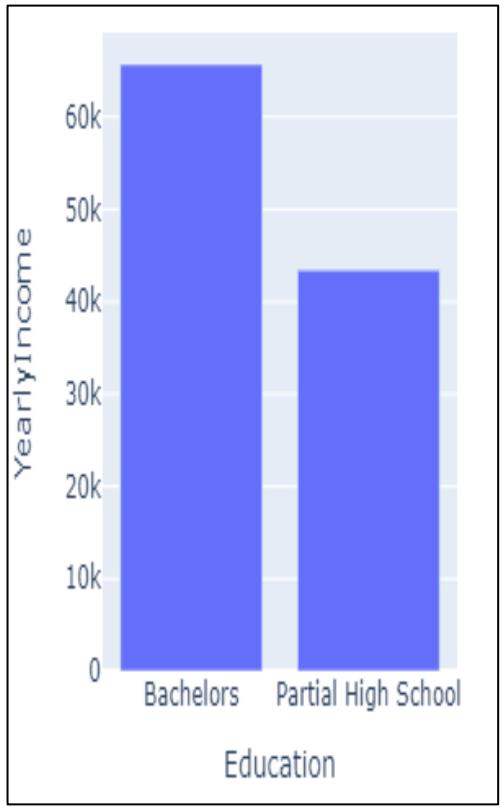
Yearly income range and purchase correlation



- High salary range leads to increase in revenue.

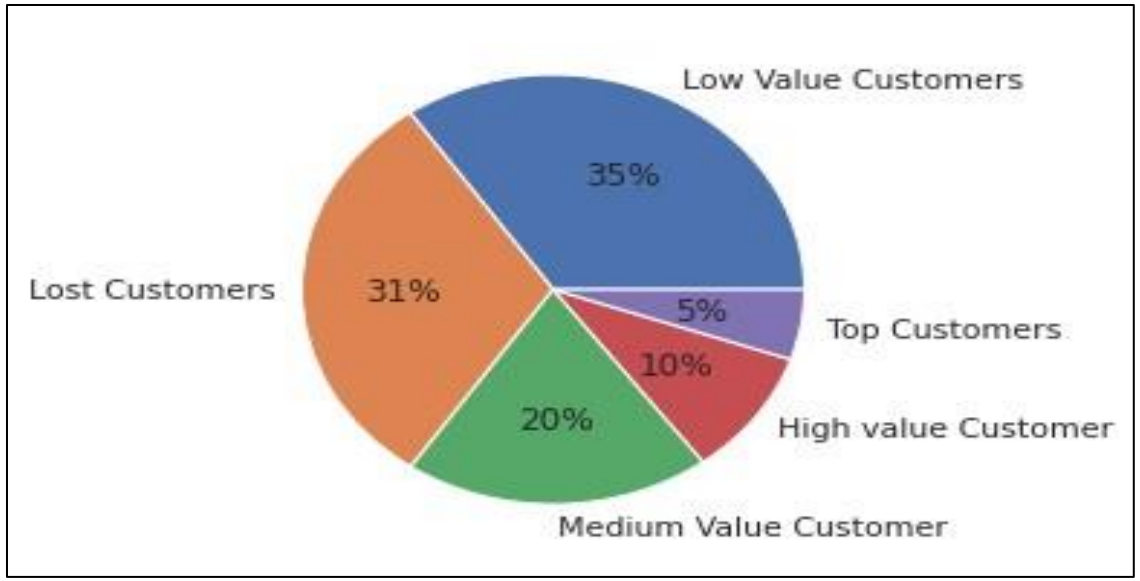
Partial high school vs bachelor's income mean and most ordered product





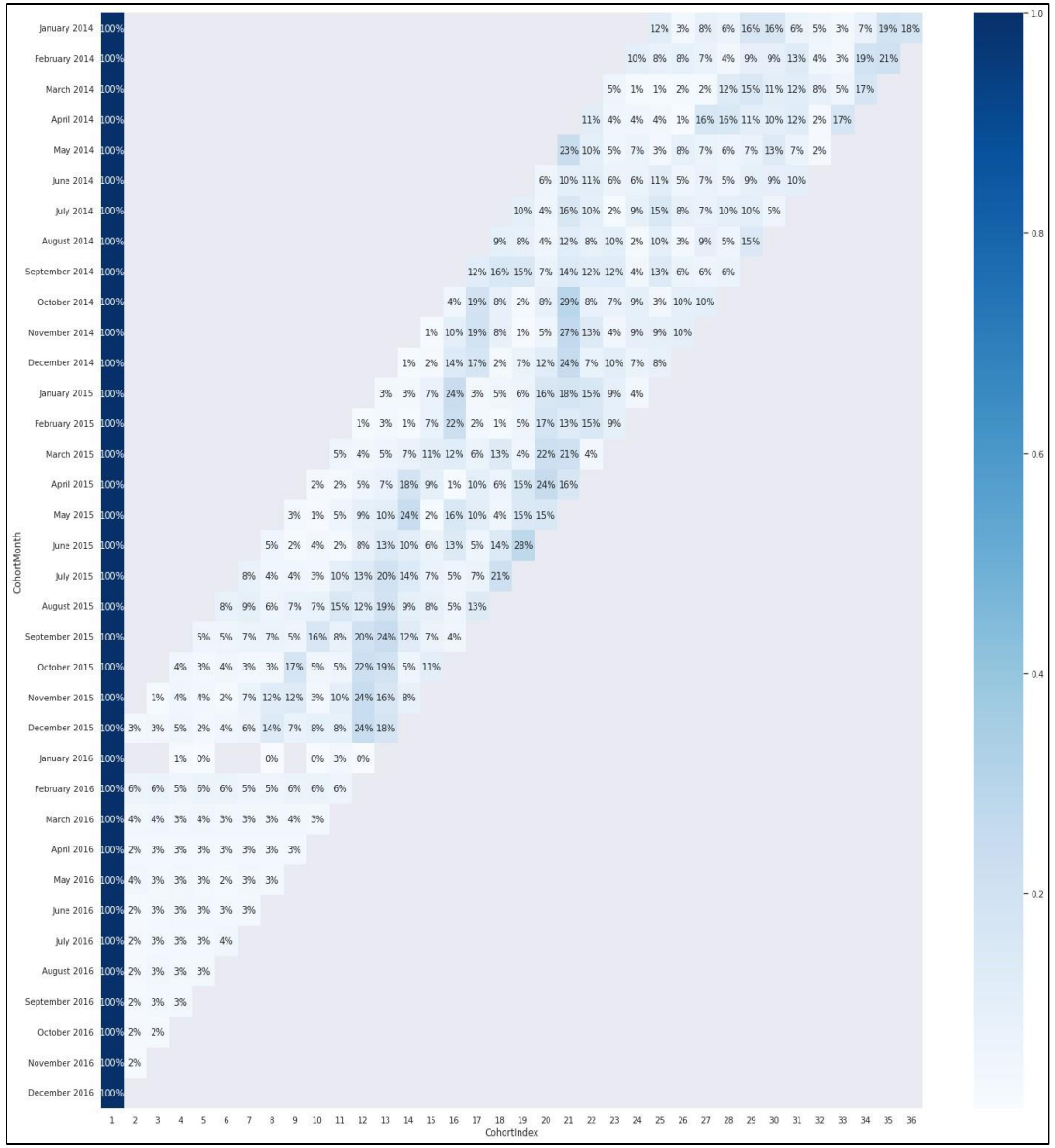
- Customers with a high school diploma and modest annual income buy more products than people with bachelor's degrees

Customer segmentation



- According to the customer segmentation described above, approximately 15% of our clients are high-value clients, whereas the majority of our clientele are low-value and lost clients.

Cohort Analysis



- We can infer from the heatmap above that client retention in 2014 was subpar.
- Since August 2015, we have noticed some customers returning, though not in large numbers.
- 2016 brought about a slight improvement in retention.

KEY PERFORMANCE INDICATOR (KPI)

1. Sales trend line
2. Cost trend line
3. Average unit cost and price
4. Revenue generated by Subcategory
5. Sales by Product Line
6. Revenue contribution by region
7. Profit contribution by region
8. Profit % by region
9. Current year profit margin vs difference in last year’s profit margin
10. Total orders
11. Total revenue
12. Variance to target comparison by category
13. Variance by month line chart
14. Actual sales and target sales matrix
15. Cohort analysis table
16. Customer retention line chart
17. Monthly spending trend
18. Average monthly spend distribution

CONCLUSION

- A sizable portion of the clientele is made up of people between the ages of 40 and 59
- The year 2016 saw an exponential surge in sales.
- High quantity of products is ordered from Australia and the United States.
- Major Profit is contributed by the Bike Category.
- The average order has a gap of 7 days between the day the order is ready for export from the factory and the date it was shipped.
- Maximum profit earned in the months of **June, November, and December.**
- High sales orders are seen on **Wednesday and Saturday** when compared to other weekdays.
- There is a high negative correlation between Price and the number of Quantities ordered.

- The average amount spent by men without permanent addresses is low, whilst the average amount spent by women without permanent addresses is higher.
- Age range of 40-49 and 50-59 shows high demand compared to other age groups.
- High salary range leads to an increase in revenue.
- Customers with a high school diploma and modest annual income buy more products than people with bachelor's degrees.
- According to the customer segmentation described above, approximately **15% of our clients are high-value clients**, whereas the **majority of our clientele are low-value and lost clients.**
- Client retention in 2014 was subpar.
- 2016 brought about a slight improvement in retention.

Q & A

Q1) What's the source of data?

- The Dataset was taken from iNeuron's Provided Project Description Document.
- Budget-Sales-Analysis/Datasets at main · Pp11112000/Budget-Sales-Analysis (github.com)

Q2) What was the type of data?

- The data was a combination of numerical and Categorical values.

Q 3) What's the complete flow you followed in this Project?

- Refer to page 5 for better Understanding.

Q4) What techniques were you using for data?

- Removing unwanted attributes.
- Visualizing relation of independent variables with each other.
- Cleaning data by removing column with missing values.
- Converting Numerical data into Categorical values.

Q 6) What were the libraries that you used in Python?

- I used Pandas, NumPy, Matplotlib, Seaborn and Plotly libraries.

THANK YOU