DETAILED PROJECT REPORT

BUDGET SALES ANALYSIS

MEMBERS:

- BANAGAR MAHESH
- MONESH S
- SANKETH P
- SIDDARTH A
- SOMANSH V

1. Problem Statement

Our "Domain Sale" process is structured to help potential buyers purchase the domain they want immediately without thehassle of contacting the seller directly.

A seller lists a domain for sale at a specific price in our Marketplace. An interested buyer sees this domain for sale and decides to buy it.

2. Objectives

- The collection includes records for sales orders, customer information, product information, and geographical data.
- In order to deduce important metrics and patterns in the dataset, this project will use the provided data to performETL and data analysis.
- Additionally, several visualisations and reports are created to represent significant linkages.

3. 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.
- Facilitates seamless resource management flow.

4. Data Attributes

Customer

CustomerKey	FullName	Birthdate
Maritalstatus	Gender	YearlyIncome
TotalChildren	NumberChildrenAtHome	Education
Occupation	HouseOwnerFlag	NumberCarsOwned
DateFirstPurchase	CommuteDistance	

Product

ProductKey	ProductName	Subcategory
Category	ListPrice	DaysToManufacture
ProductLine	ModelName	ProductDescription
StartDate		

Territory

SalesTerritoryKey	Region	Country
Group		

Sales

ProductKey	OrderDate	ShipDate
CustomerKey	PromotionKey	SalesTerritoryKey
SalesOrderNumber	SalesOrderLineNumber	OrderQuantity
UnitPrice	TotalProductCost	SalesAmount
TaxAmt		

4.1 Dataset Information

CustomerKey: Primary key for customer dataset

Birthdate: Birthdate of the customer

MaritalStatus: M- Married / S - Single

Gender: M – Male / F – Female

TotalChildren: Total number of children

NumberChildrenAtHome: Number of children staying along withtheir parents

Education: Education qualification

Occupation: Present occupation

HouseOwnerFlag: 1– Owns house / 0- Doesn't have a permanent address

NumberCarsOwned: Number of cars owned by the customer

DateFirstPurchase: First date of order by the customer

ProductKey: Primary Key for the product dataset

ProductName: Product name with colour of the product

Subcategory: Sub category name of the product

Category: Category name of the product

ListPrice: Sale price of the product

DaysToManufacture: Days to manufacture the product afterreceiving the order

ProductLine: Product line name

ModelName: Model name of the product

ProductDescription: more details about the product

SalesTerritoryKey: Primary Key of the Territory dataset

Region: Region name of the order

Country: Country name of the order

OrderDate: Date of the order received

ShipDate: Date when the order left the factory for export

SalesOrderNumber: Invoice number of the order

OrderQuantity: Number of quantities ordered for a product

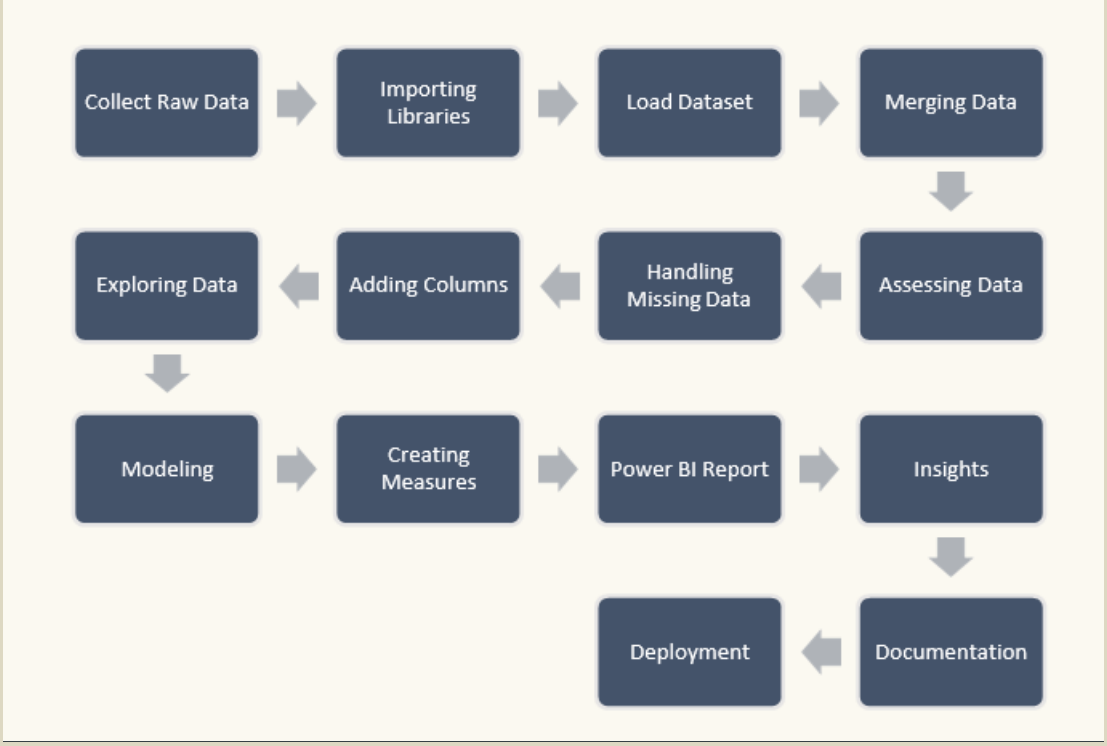
UnitPrice: Per unit sale price of the product

TotalProductCost: Cost of the product

SalesAmount: Total sales price of the product

TaxAmt: Tax collected for the product sold

5. Architecture



1. **Collect Raw Data -** This step involves extracting the data from different sources relevant to the problem statementor obtaining data from the client

2. **Importing Libraries –** Import analysis related python libraries example – Pandas, Numpy, Plotly, datetime etc

3. **Data Wrangling –** Contains following steps gathering data, assessing data, handling missing data and adding column

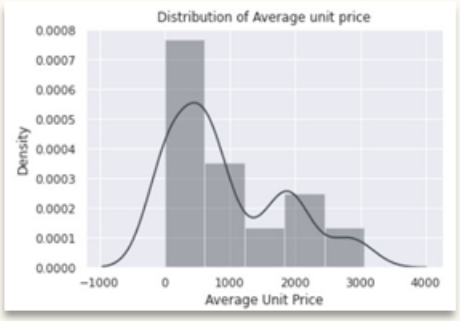
4. **Exploring Data –** Once the data is loaded and pre- processed, we preform data analysis using python librariesand Business Intelligence tools like Power BI

5. **Data Modelling** - Data Modelling is one of the features usedto connect multiple data sources in BI tool using a relationship. A relationship defines how data sources are connectedwith each other and you can create interesting data visualizations on multiple data sources

6. **Deployment -** The prepared visualizations are deployed onthe powerbi.microsoft.com site. Where they will be available publicly

6.Insights

1. Product Price per unit Distribution



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

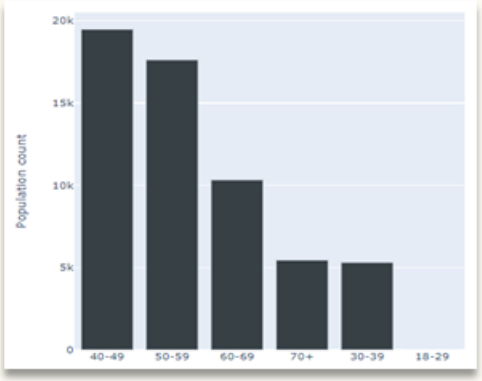
2. Sales order line number distribution

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

3. Sales order quantity distribution

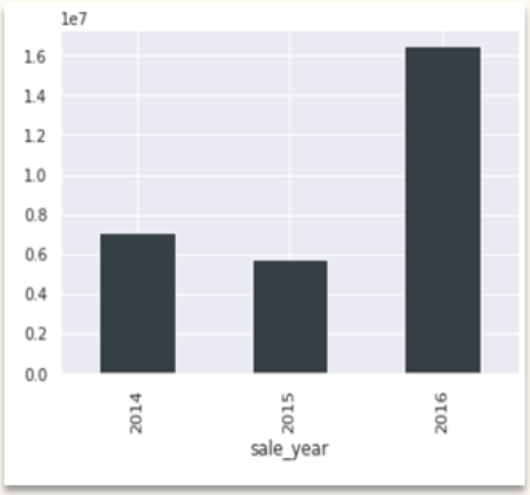
- Maximum quantity ordered for a product is below 5

4. Age distribution



- A sizable portion of the clientele is made up of people between the ages of 40 and 59

5. Year wise sales



- The year 2016 saw an exponential surge in sales

7. 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
	Accessories	Bike Racks	493
		Bike Stands	394
		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

9. **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
	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
2016	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

- Major Profit is contributed by the Bike Category

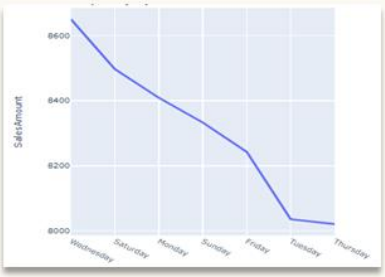
10. **How efficient are the logistics?**

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

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

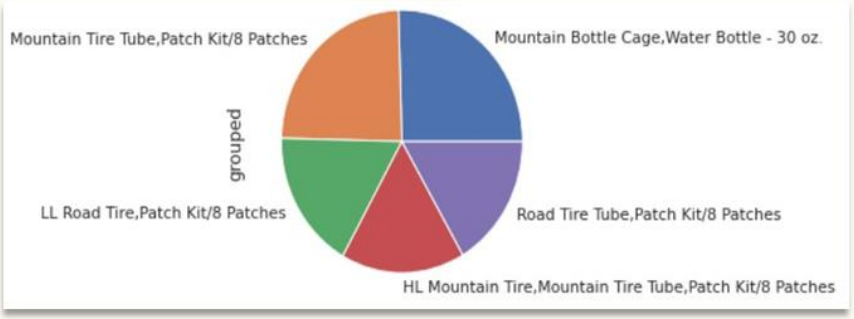
- Maximum profit earned in the months of **June, November, and December**

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



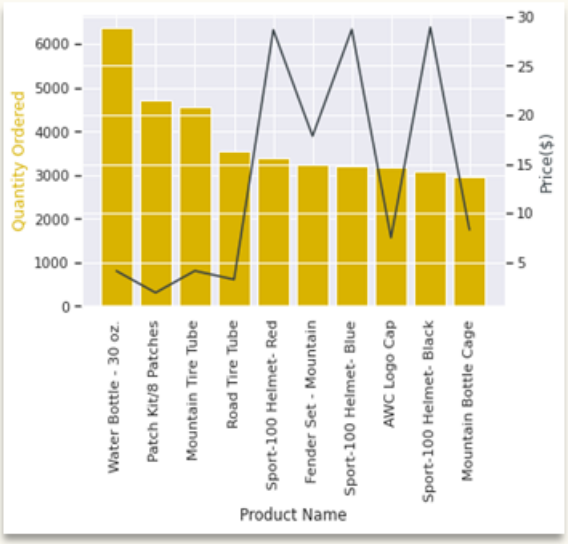
- High sales orders are seen on **Wednesday and Saturday**; therefore, we can promote our product during these workweek

13. **Which products are most often sold together?**



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

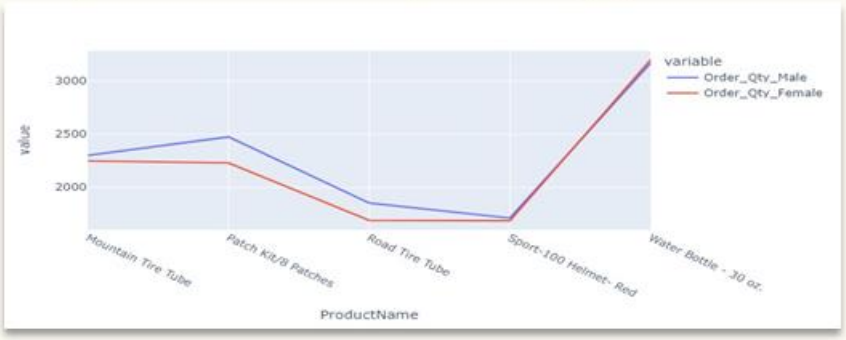
14. **Which product sold the most? why do you think it sold the most?**



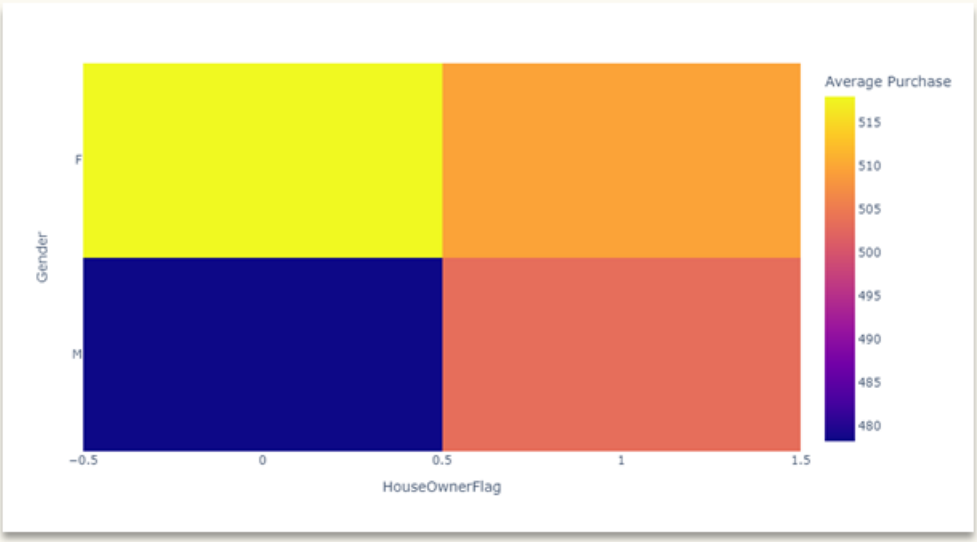

```
prices.corr(quantity_ordered)
-0.5333019792658484
```

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

15. **Compare most ordered product by gender**

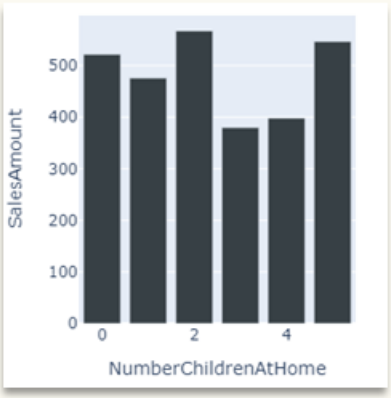


16. **Does Gender and home ownership matter in order purchasing**



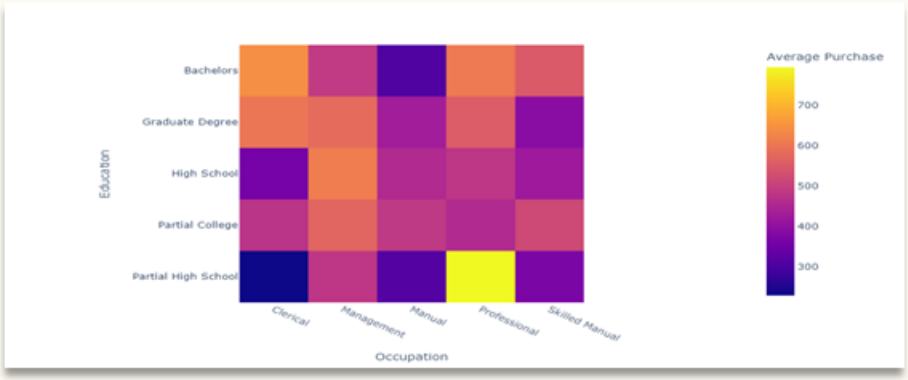
- It's interesting to note that the average amount spent by men without permanent addresses is low, whilst the average amount spent by women without permanent addresses is higher

17. Number of children and Purchase correlation



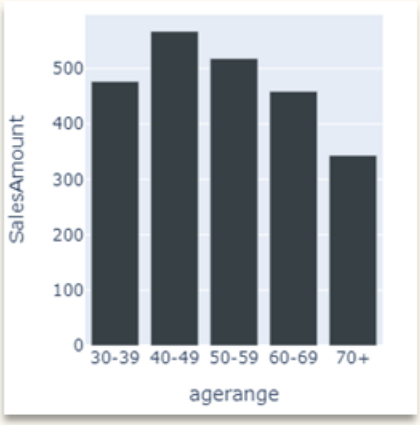
- Purchase among customers with number of children, 2 and 5, are high

18. Occupation and purchase correlation



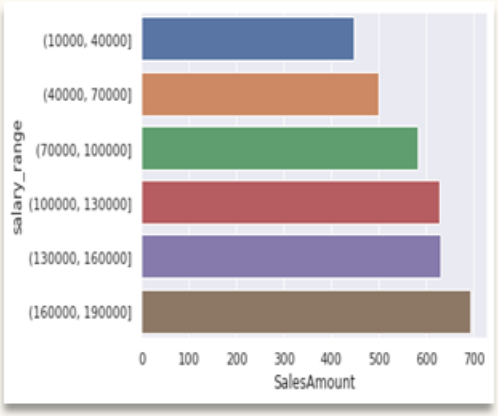
- Purchases by Professional and Management customers are comparatively high

19. Which age group has produced the most revenue?



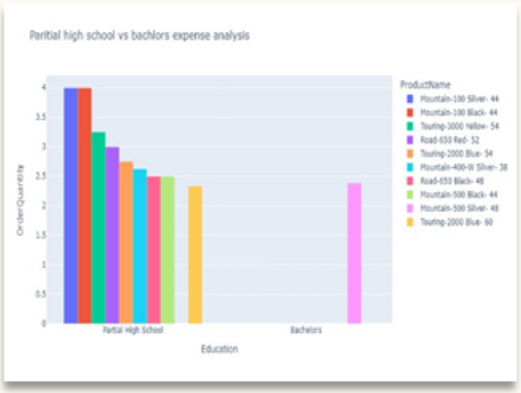
- Age range of 40-49 and 50-59 is shows high demand compared to other age group

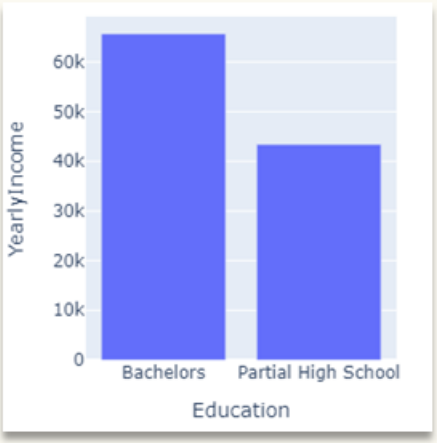
20. Yearly income range and purchase correlation



- High salary range leads to increase in revenue

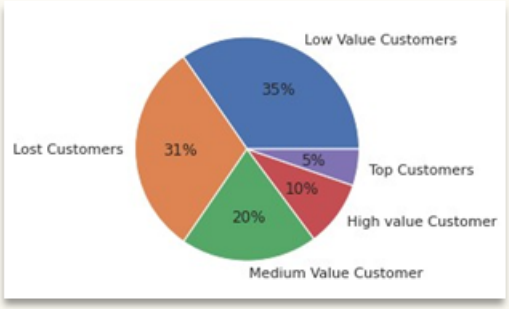
21. Partial high school vs bachelors income mean and most ordered product



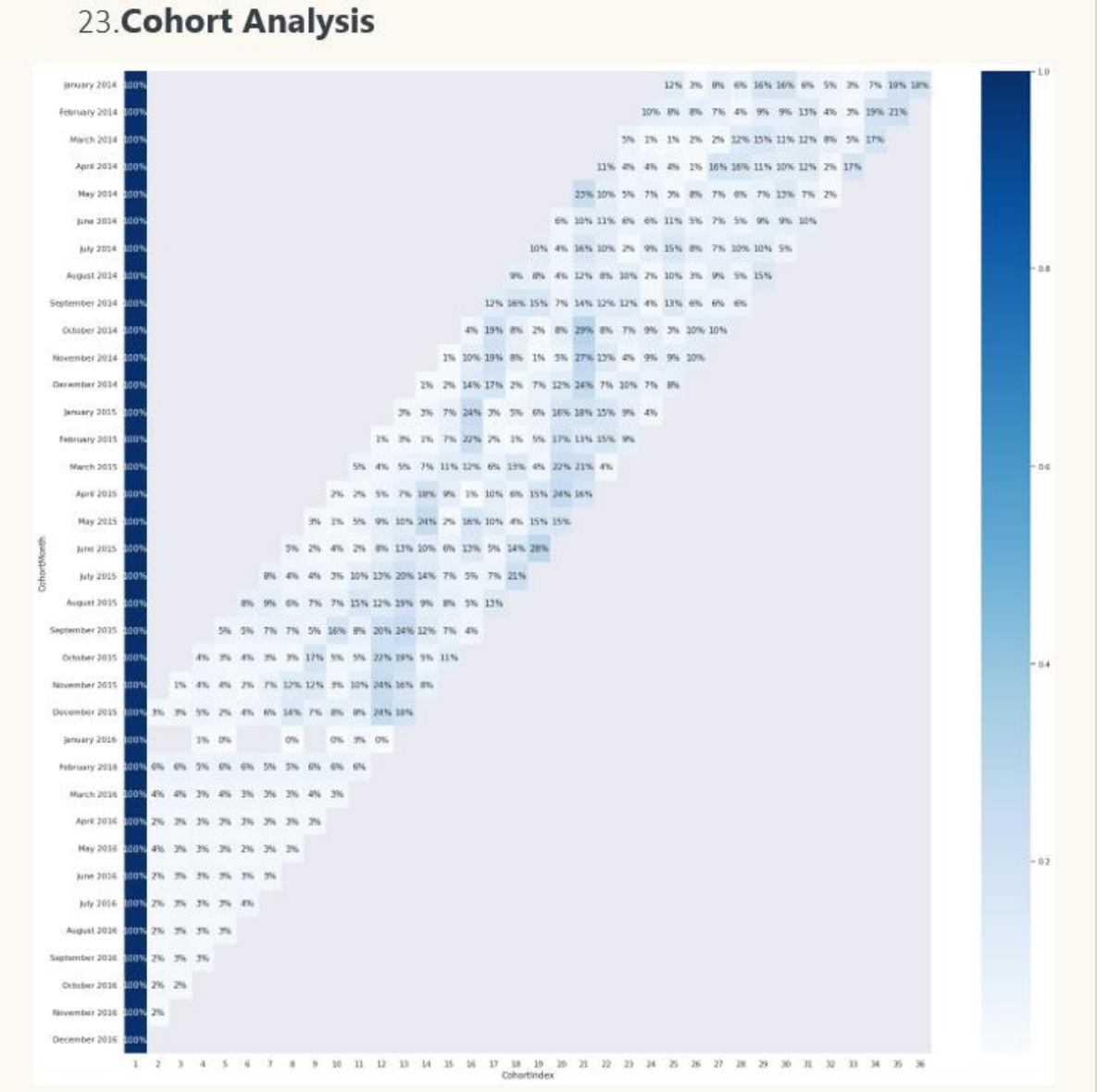


- Customers with a high school diploma and modest annual income buy more products than people with bachelor's degrees

22. 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**



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

7. Key Performance Indicators

- Sales trend line
- Cost trend line
- Average unit cost and price
- Revenue generated by Subcategory
- Sales by Product Line
- Revenue contribution by region
- Profit contribution by region
- Profit % by region
- Current year profit margin vs difference in last year's profit margin
- Total orders
- Total revenue
- Variance to target comparison by category
- Variance by month line chart
- Actual sales and target sales matrix
- Cohort analysis table
- Customer retention line chart
- Monthly spending trend
- Average monthly spend distribution

8. 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 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 number of Quantity 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 is shows high demand comparedto other age group
- High salary range leads to 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

9.Q & A section

Q1) What's the source of data?

- The Dataset was taken from iNeuron's Provided Project Description Document
- Data Link

Q2) What was the type of data?

- The data was the combination of numerical and Categorical values

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

- Refer page 4 for better Understandings

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