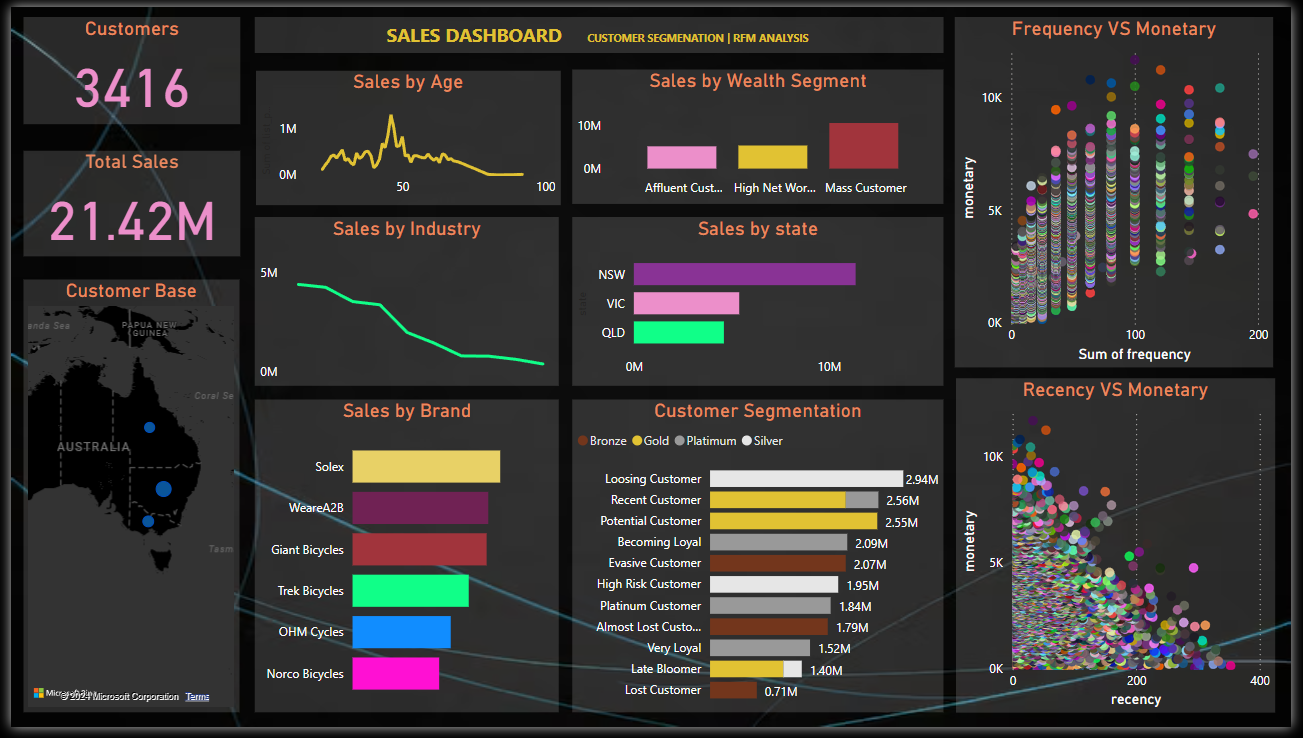
# Data Analytics Customer Segmentation

## Goal of the project

The goal of this project is to perform a Customer Segmentation Analysis for an automobile bike company using a Recency, Frequency, Monetary (RFM) model. RFM analysis is a behavior-based approach that segments customers based on their past purchase transactions. In this study, customers were divided into 11 segments. This analysis will identify which customer segments should be targeted to increase sales revenue for the company. A sales dashboard for customer segmentation is developed using **Power BI**, while data quality assessment and analysis are conducted using **Python**.

## Power Bi Dashboard



## Analysis Approach

### 1. Data Quality Assessment and Data Cleaning

The initial step in generating useful insights from the data involved data preparation, quality assessment, and data cleaning. Following the cleaning process, exploratory data analysis was conducted to identify customer purchasing behaviors and generate insights.

During the data cleaning phase, the quality of the following datasets was assessed, revealing several data quality issues which were subsequently addressed:

**CustomerDemographics.xlsx:**

1. An irrelevant column was identified and removed from the dataset.

2. Five columns contained missing values. Depending on the volume of missing data, records were either dropped or appropriate values were imputed.

3. The gender column lacked standardization, so the data was standardized to remove inconsistencies.

4. The Date of Birth column was transformed to create new feature columns, 'Age' and 'Age Group,' to check for age distribution discrepancies. An outlier was identified and removed.

5. No duplicate records were found in this dataset.

**NewCustomerList.xlsx:**

1. Five irrelevant columns were identified and removed from the dataset.

2. Four columns contained missing values. Depending on the volume of missing data, records were either dropped or appropriate values were imputed.

3. The Date of Birth column was transformed to create new feature columns, 'Age' and 'Age Group.'

4. No data inconsistencies were found.

5. No duplicate records were found in this dataset.

**Transaction\_data.xlsx:**

1. The product\_first\_sold\_date column was not in datetime format; it was converted from int64 to datetime format.

2. Seven columns contained missing values. Depending on the volume of missing data, records were either dropped or appropriate values were imputed.

3. A new feature column, 'Profit,' was created, representing the difference between the list price and the standard price.

4. No data inconsistencies were found.

5. No duplicate records were found in this dataset.

**CustomerAddress.xlsx:**

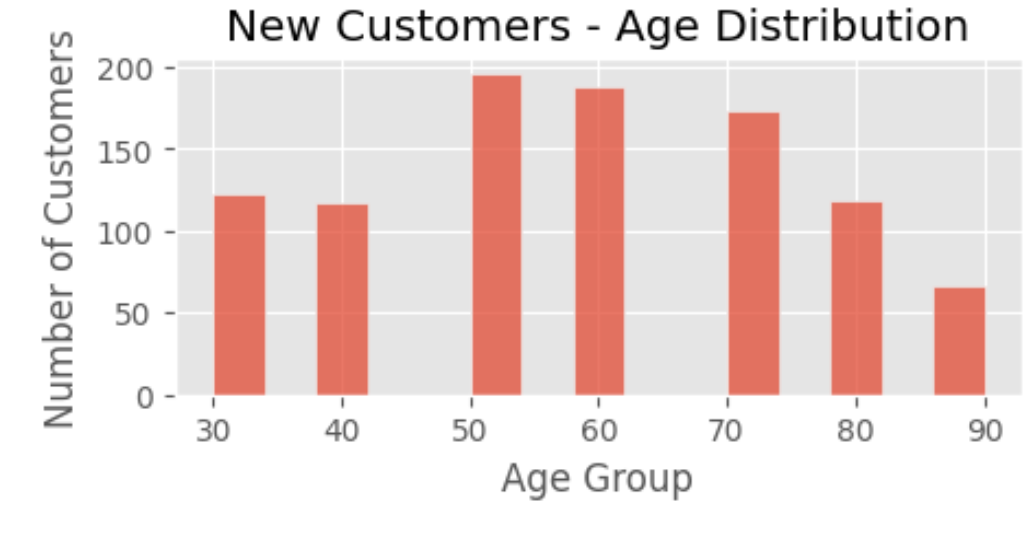
1. The states column lacked standardization, so the data was standardized to remove inconsistencies.

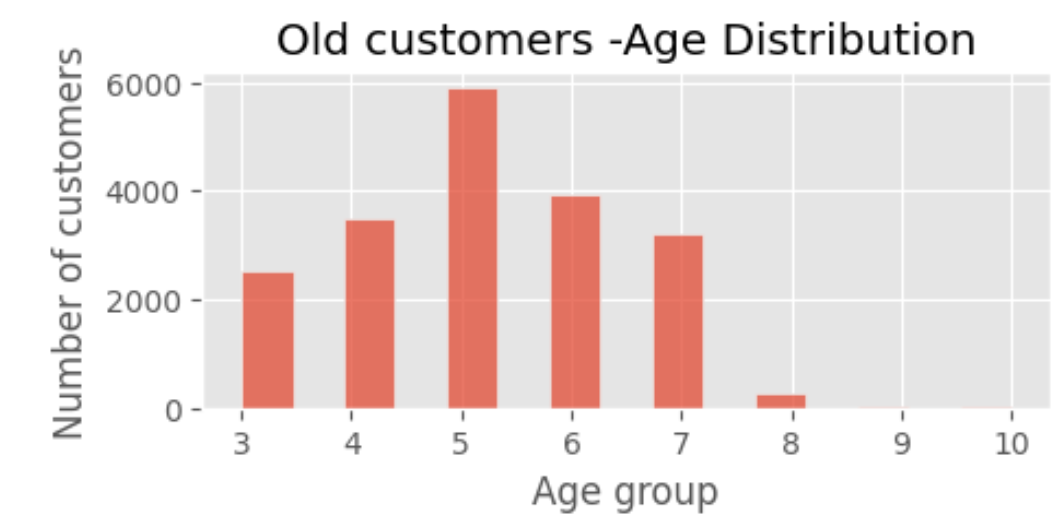
2. Some customer IDs from the Customer Demographics table were missing in the Address table, leading to dropped records.

### 2. Exploratory Data Analysis on Customer Segments

After the data cleaning process, exploratory analysis on the dataset was performed, yielding the following insights:

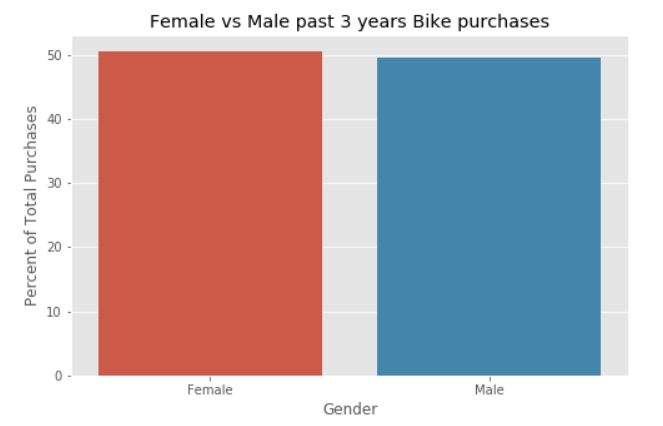
**New vs. Old Customers Age Distribution:**

* The majority of both new and old customers are aged between 40-49.
* The lowest number of customers for both groups falls in the under-20 and over-80 age brackets.
* The automobile company is particularly popular among new customers in the 20-29 and 40-49 age groups.
* A significant decline in new customers is observed in the 30-39 age group.



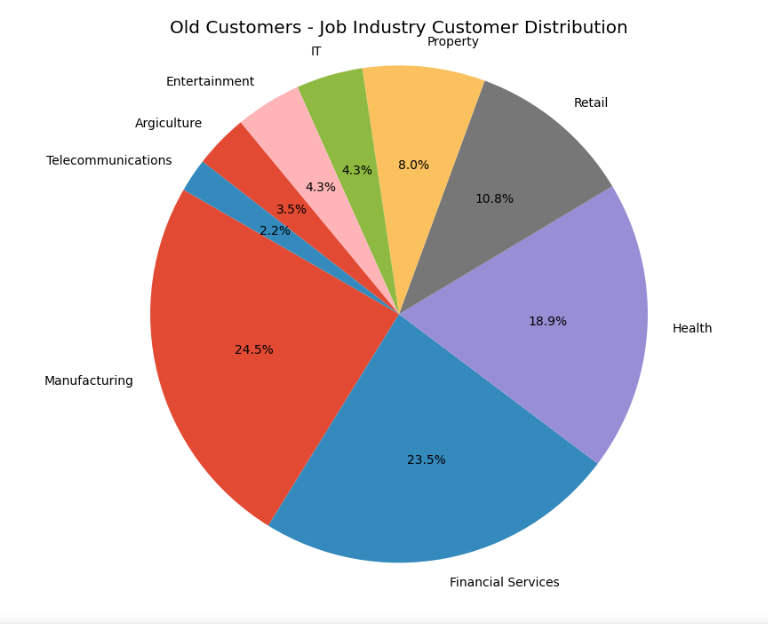
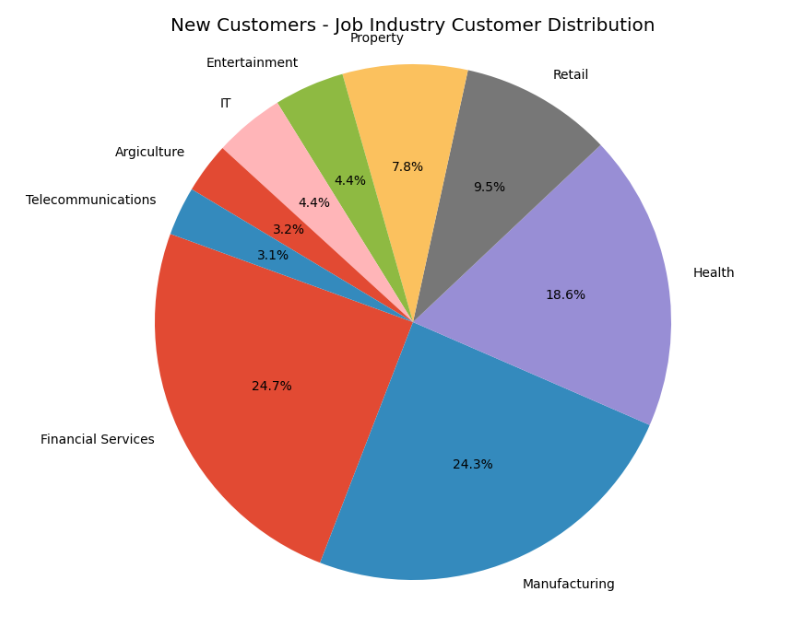
**Bike purchases over last 3 years by Gender**

* Most bike purchases are done by Female over the last 3 years. Approximately 51% of the bike purchases are done by Female compared to 49% of the purchases being done by Male.
* The Female purchases are 10,000 more than that of Male purchases (numerically).

[](https://github.com/AbhishekGit-hash/Data-Analytics-Customer-Segmentation/blob/master/data%20visualization/Female%20vs%20Male%20Bike%20Purchases.PNG)

**New vs Old Customers Job Industry Distribution**

* Most New customers are from the Manufacturing and Financial Services sector (approx. 20% of the New Customers).
* The lowest number of customers are from the Agriculture and Telecom sector approx. 3%.
* Similar trend is observed among Old Customers as well.

  
**Wealth Segmentation by Age Category**

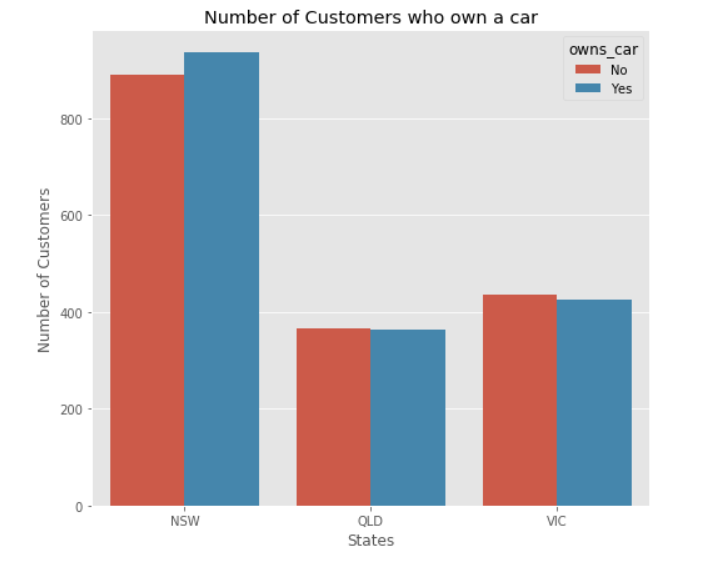
* Across all age categories the largest number of customers are from 'Mass Customer' Segment
* The next category comes from the 'High Net Worth' customers.

In the age group 40-49, Affluent segment out performs the High Net Worth customers in terms of number of customers.

|  |  |  |
| --- | --- | --- |
| **Old Customers Wealth by Age Group** |  | **New Customers Wealth by Age Group** |
|  |  |  |

**Cars owned by States**

* New South Wales has the largest number of people who do not own a car.
* In Victoria the proportion is quite even.
* In Queensland the number of people owning a car is greater than who do not have a car.

[](https://github.com/AbhishekGit-hash/Data-Analytics-Customer-Segmentation/blob/master/data%20visualization/Car%20Owners%20by%20State.PNG)

### 3. RFM Analysis and Customer Segmentation

In this stage of the analysis, customer segmentation was performed using an RFM Model. RFM (Recency, Frequency, Monetary) analysis is a behavior-based approach that groups customers based on their previous purchase transactions.

For this analysis, customers were divided into 11 segments:

Platinum Customers

Very Loyal Customers

Recent Customers

Potential Customers

Lost Customers

Losing Customers

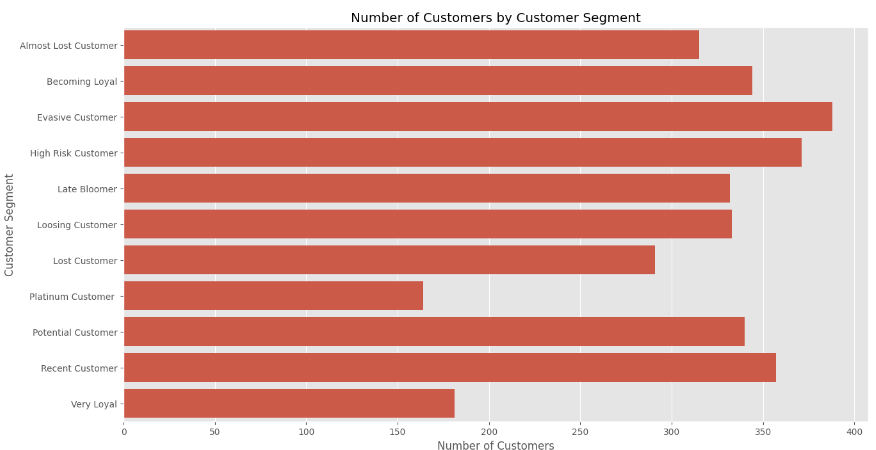
Late Bloomers

High-Risk Customers

Evasive Customers

Becoming Loyal

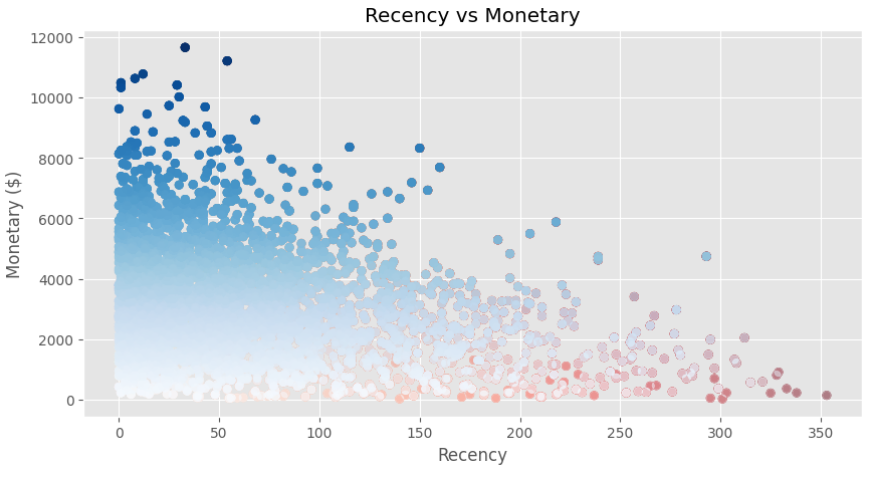
Almost Lost Customers The current distribution of customer segments in the automobile business is illustrated below:



### 4. RFM Analysis: Scatter Plots

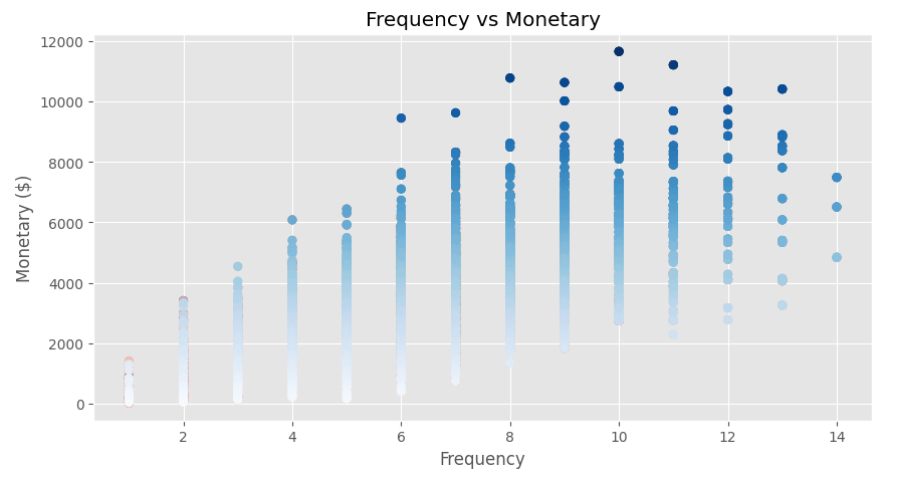
#### Recency vs Monetary:

The visualization shows that recent customers have purchased more products and generated relatively more revenue than the customers who visited a while ago.



#### Frequency vs Monetary:

The visualization shows that customers belonging to Platinum/ Very Loyal/ Becoming Loyal Customer Segments have a greater frequency and generate greater monetary for the business



## Datasets Used

The datasets used include:

**Raw\_data.xlsx**: This excel file dataset included the following sheets of data:

* + **Transactions\_data.xlsx**: This dataset included the transactions data of the customers across all the different states in Australia.
  + **NewCustomerList.xlsx**: This dataset included the new customers who visted the automobile bike company recently.
  + **CustomerDemographic.xlsx**: This dataset included entire details of the Customer Demographics.
  + **CustomerAddress.xlsx**: This dataset included the address of the Customers.

## Tools and Technologies used

The tools used in this project include:

* **Python** - This was needed to conduct **Data Quality Assessment** and also for **Data Cleaning processes**. With Python libraries **pandas, matplotlib, seaborn** exploratory data analysis of the datasets and to gain useful insights from the data was possible.
* **Power Bi** - This **Business Intelligence** tool was required to explore data and create charts, graphs, visualizations to come up with a **Sales Dashboard for Customer Segmentation** for the automobile bike company. The Power Bi Sales Dashboard can be found

## Built With

- Python 3.8.2, Power Bi

## Authors

* Dhir Luthra – [GitHub Profile](https://github.com/dhirluthra19)

|  |  |
| --- | --- |
|  | |
|  |  |