**WhaleCo Streaming Analysis**

**Introduction**

WhaleCo, a leading TV streaming service provider, aims to optimize its offerings and improve customer satisfaction. This analysis focuses on WhaleCo's customer base and streaming packages using the provided datasets: "WhaleCo Members" and "WhaleCo Orders."

**Objectives**

1. *Understand WhaleCo's customer demographics*, including distribution across different states, occupations, and salary ranges.
2. *Evaluate the performance of WhaleCo's streaming packages*, identify popular packages, assess customer satisfaction levels, and analyse the relationship between package attributes and customer behavior.
3. *Segment WhaleCo's customers* based on their subscription tiers (Bronze\_Member, Silver\_Member, Priority\_Member) and identify patterns or differences in their preferences and behavior.
4. *Conduct churn analysis* to calculate the churn rate and identify factors contributing to customer churn.

The "WhaleCo Members" dataset contains details such as customer ID, address, occupation, salary, subscription date, and active months. The "WhaleCo Orders" dataset includes information about the streaming packages ordered by customers, including the package name, price, customer satisfaction scores (OSAT and NPS), streaming activity, and the number of internal accounts associated with each order.

**Methodology**

**1. Data Preprocessing:**

* **Python:** Cleaned the dataset, addressing missing values, inconsistencies, and data type conversions, ensuring data integrity and accuracy. Leveraging pandas and numpy libraries, I transformed categorical variables into numerical format for meaningful analysis. By merging multiple datasets, I created a unified dataset, enhancing data consistency and enabling comprehensive insights.
* **Power BI:** Imported and meticulously transformed the datasets, ensuring data integrity and consistency. By merging relevant datasets, I established meaningful relationships, laying the foundation for in-depth analysis and insightful visualizations.

**2. Exploratory Data Analysis (EDA):**

* **Python:** Conducted EDA to understand data distribution, identify trends, and derive insights using matplotlib and seaborn for visualizations.
* Crafted visualizations that not only captured data distributions but also revealed compelling trends and invaluable insights.
* **Power BI:** Data Modelling, created model view, data view and visualizations (charts, graphs) to explore customer demographics, product performance, customer segmentation, and churn analysis.

**3. Metrics Creation:**

* **Python:** Calculated metrics such as churn rate, customer satisfaction scores, and other relevant measures using pandas.
* **Power BI:** Created new columns Churn status, Churn rate columns by using DAX function and measures to derive churned insights.
* Churn Status = IF(NOT(ISBLANK([unsub\_date])), "Active", "Churned")
* Churn Rate = DIVIDE (

    CALCULATE(COUNTROWS(Customers), Customers [Churn Status] = "Churned"),

    COUNTROWS(Customers))

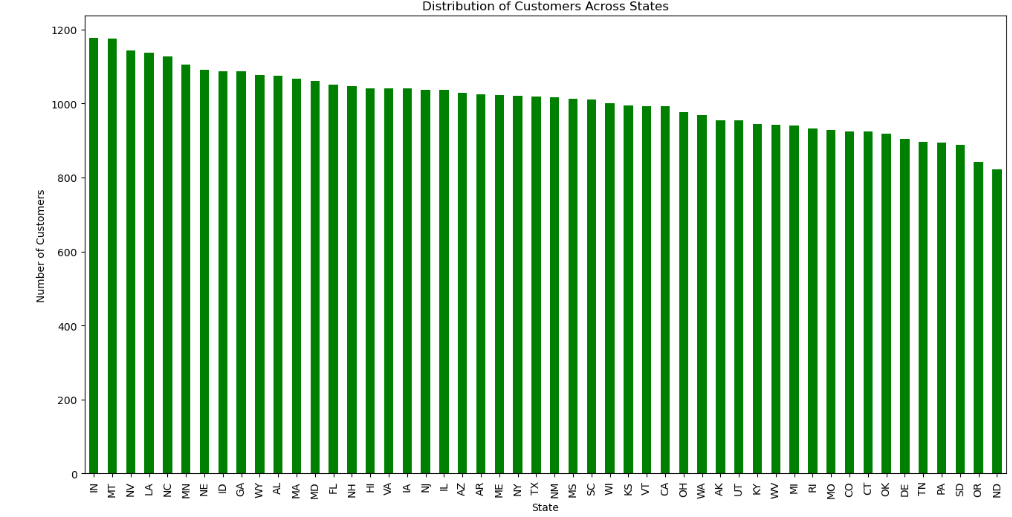
**4. Machine Learning Models (Python):**

Implemented machine learning model (Logistic Regression) for, customer segmentation, and customer behaviour prediction using Scikit- learn. and checked accuracy of model, along with the Confusion Matrix.

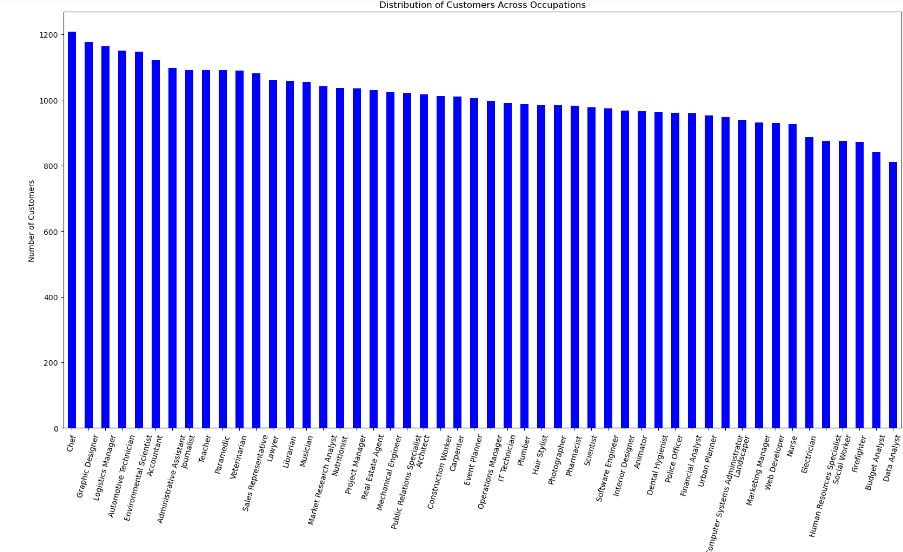
**Results**

The EDA revealed key insights:

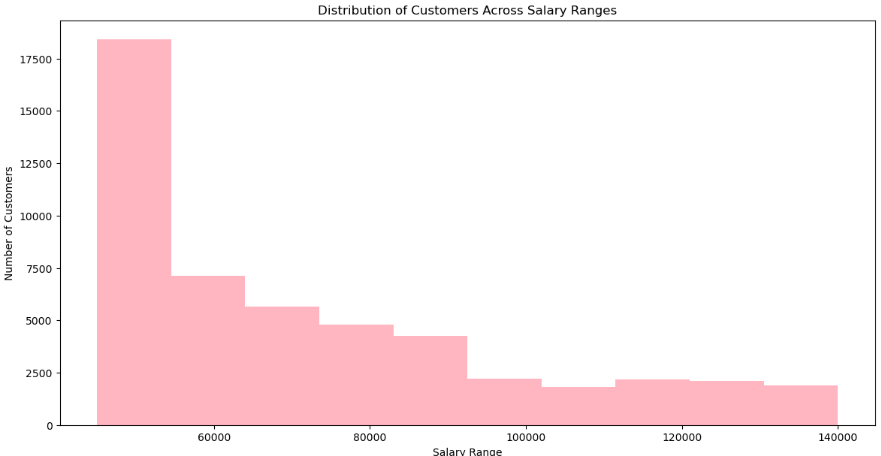
**1. Customer Demographics:**



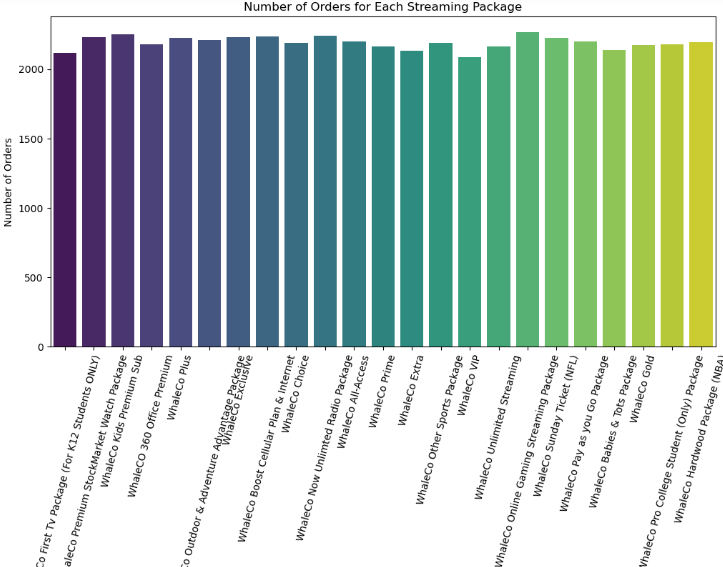
*Observation:* Majority of customers are located in urban areas, with ‘IN’ having the highest customer density.



*Observation:* Distribution of customers across various occupations indicates a diverse user base.

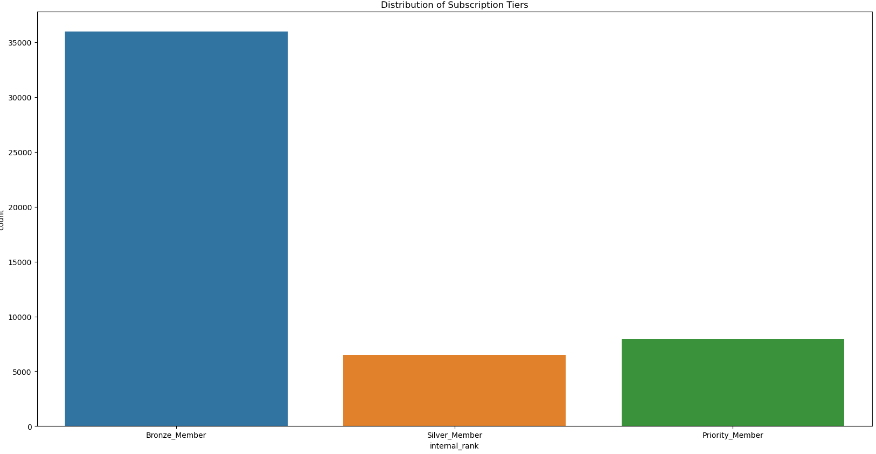


*Observation:* Customers salary ranges vary, with a significant portion falling within the middle-income bracket.

**2. Product Performance:**

*Observation:* WhaleCo Unlimited Streaming and WhaleCo kids premium sub are the most popular packages, with high engagement and satisfaction scores.

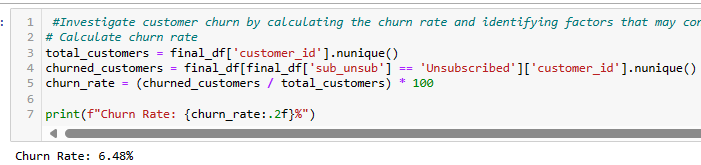
**3. Customer Segmentation:**

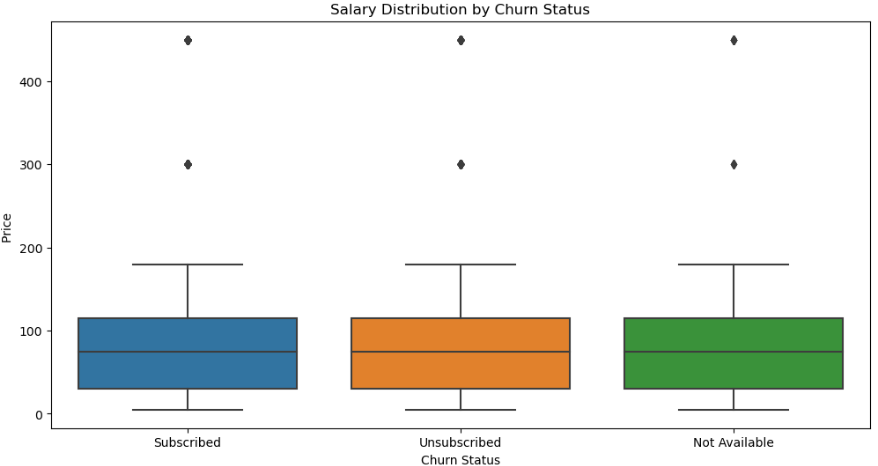


*Observation:* Three main subscription tiers Bronze \_Member, Silver\_Member, Priority\_Member are identified, with distinct preferences and behaviours.

Bronze Members has largest segment, followed by Priority Members and Silver Members.

Preferences and behaviour differ among segments, with Bronze Members showing higher engagement and satisfaction levels.

**4. Churn Analysis:**



*Observation:* Churn rate analysis indicates a moderate churn rate, with factors such as Customer behaviour and pricing being potential contributors to churn.

**Machine Learning Result**

Accuracy**: 0.92**84864114263043

*Observation:* The logistic regression model shows a promising 93% accuracy in predicting whether customers are subscribed. This aligns with our goal of understanding customer behavior. While it performs well for 'Subscribed' customers, it could be improved for predicting 'Not Available' and 'Unsubscribed' categories, which are crucial for our retention strategies.

**Recommendations**

**Streaming Packages Improvement:**

* Enhance content offerings and personalize packages to cater to diverse customer preferences.
* Introduce flexible pricing options or bundle deals to attract new customers and retain existing ones.

**Customer Satisfaction Enhancement:**

* Invest in improving streaming quality and reliability to enhance overall customer experience.
* Implement customer feedback mechanisms to continuously monitor and address customer concerns promptly.

**Retention Strategies:**

* Develop targeted retention campaigns based on customer segmentation to prevent churn.
* Introduce loyalty programs or rewards for long-term subscribers to incentivize continued engagement.

**Conclusion**

* This analysis provides valuable insights into WhaleCo's customer demographics, product performance, customer segmentation, and churn analysis. By implementing the recommended strategies, WhaleCo can improve customer satisfaction, retention, and overall business performance, thereby solidifying its position in the competitive TV streaming market.