

E-COMMERCE CUSTOMER BEHAVIOUR AND SALES ANALYSIS

End-to-End Data Analytics Using Python,SQL & Power Bi

NAME:Papu Singh Deo

Problem Statement

The business aims to understand customer behavior and revenue drivers in order to improve retention and operational decision-making. Using historical e-commerce data, the objective is to analyze customer segmentation, demand patterns, pricing and discount effectiveness, delivery performance, and churn risk signals to identify actionable insights that can help reduce customer attrition, optimize marketing and pricing strategies, and enhance overall customer experience.

Dataset Overview

SOURCE : E_Commerce Dataset(kaggle)

Tables Used : Customer Behaviour table

Tools Used : Python,SQL and Power Bi

Project WorkFlow

Data Extraction

Time series Analysis

Data Cleaning

Sql Analytics

EDA

Hypothesis Testing

Power Bi Dashboard

Data Cleaning and Processing

- Removed Null Customer Values
- Remove negative quantity rows
- Created Age_Bucket and City_Tier
- Transform Coumns into proper format

Customer Segmentation Insights

Adults and Mid-Senior customers dominate the customer base and revenue, while Seniors contribute a relatively small share.

Tier-3 cities contribute the highest revenue across all age groups, driven by higher customer volume

Mid-Senior customers have the highest percentage of returning customers, indicating stronger loyalty.

Age alone is not a strong driver of spending behavior, as average spending differences across age groups are statistically insignificant.

Engagement Behavior Insights

Session duration and pages viewed are broadly similar across age groups, suggesting comparable browsing behavior.

Engagement metrics alone do not clearly differentiate high-value customers by age.

Product & Demand Insights

Product preferences vary by age group, but core categories (Food, Sports, Electronics, Fashion) are consistently popular.

Tier-3 cities show stronger demand across most product categories, reinforcing their importance for growth.

Revenue & Time Series Insights

Overall revenue shows fluctuations with a stable long-term trend, indicating steady business performance rather than sharp growth or decline.

Monthly revenue exhibits seasonality, with certain months consistently outperforming others.

Rolling average analysis smooths daily volatility and suggests no strong short-term downturn, indicating stable near-term sales direction.

Tier-3 cities consistently outperform Tier-1 and Tier-2 in revenue over time.

Pricing & Discount Insights

Average Order Value (AOV) is similar for discounted and non-discounted orders, indicating discounts do not increase basket value.

Quantity purchased does not significantly increase with discounts, suggesting discounts are not effective in driving higher volumes.

Discounts perform similarly for new and returning customers, showing limited differential impact.

Statistical (Hypothesis Testing) Insights

Discounts do not lead to a statistically significant increase in quantity purchased.

No statistically significant difference in spending across age groups.

Delivery time does not have a statistically significant relationship with customer ratings, implying customers may tolerate delivery delays or value other factors more.

SQL Base Insights

Customer churn risk is high, indicating retention is a bigger challenge than customer acquisition for the business.

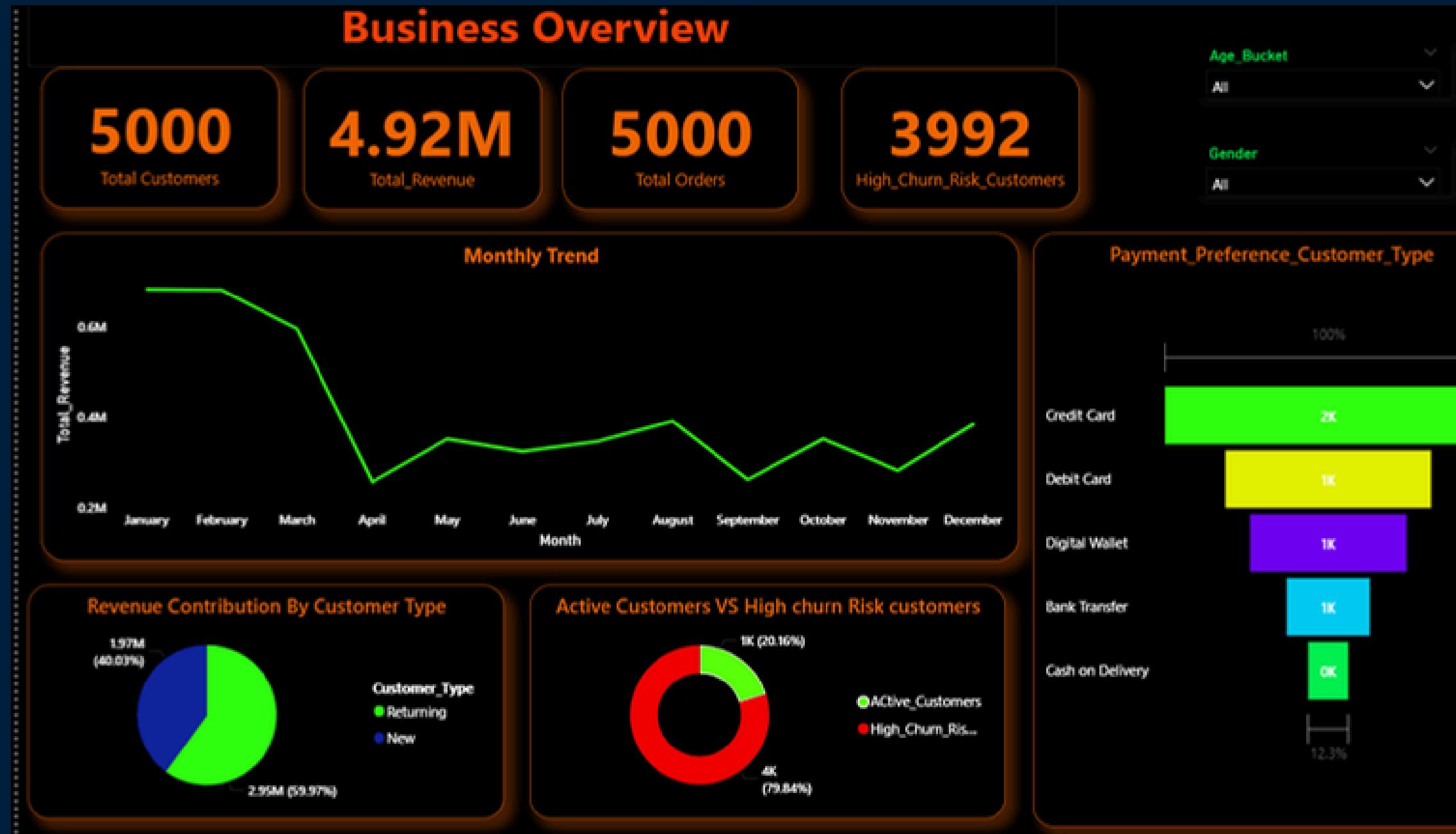
Returning customers contribute significantly more revenue than one-time buyers, making loyalty and retention strategies critical.

Discounts do not meaningfully increase order value or purchase quantity, suggesting blanket discounting is ineffective.

Revenue and demand are concentrated in specific cities and product categories, highlighting the importance of city- and category-focused strategies.

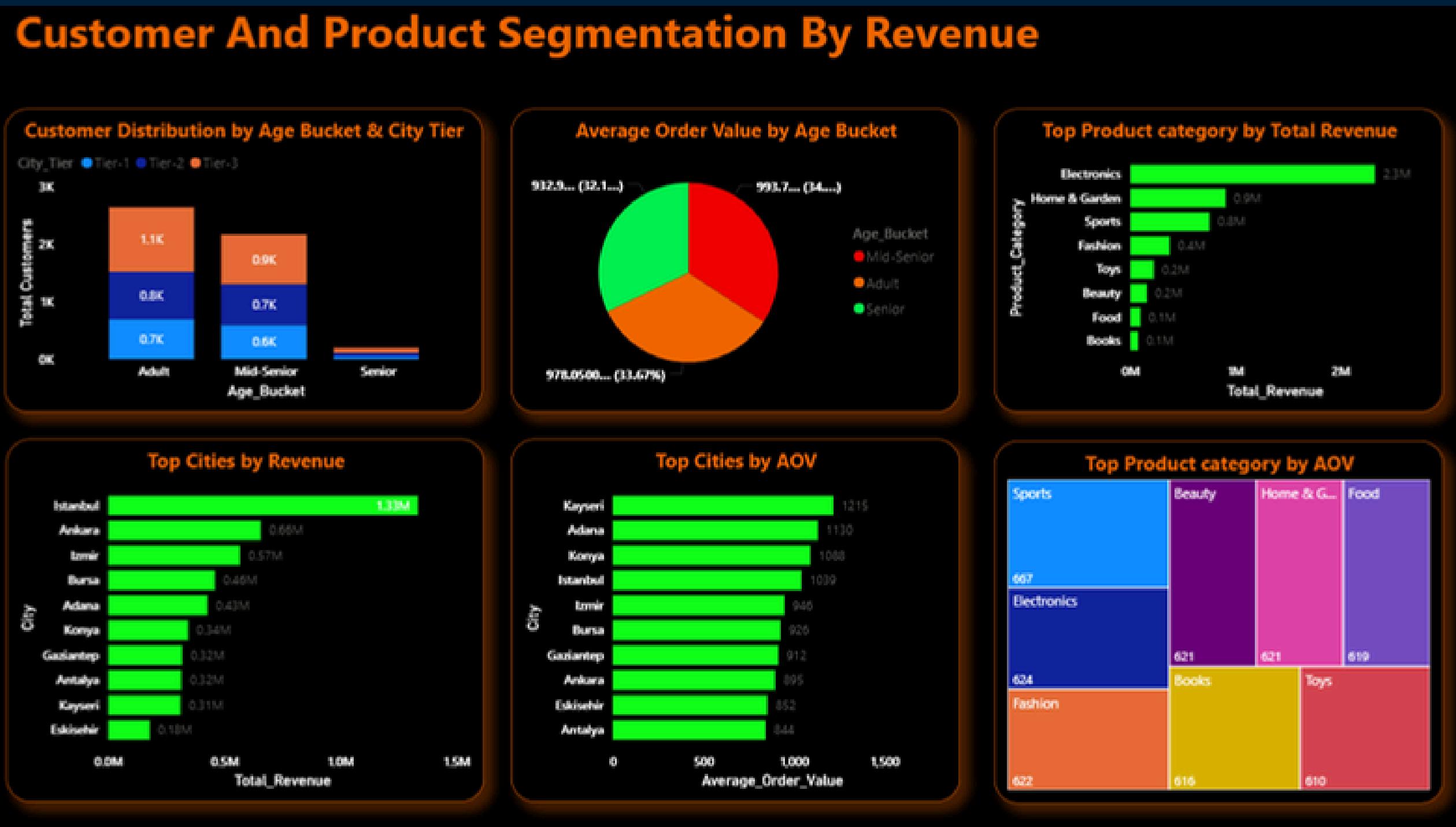
Power BI Dashboard

Page 1



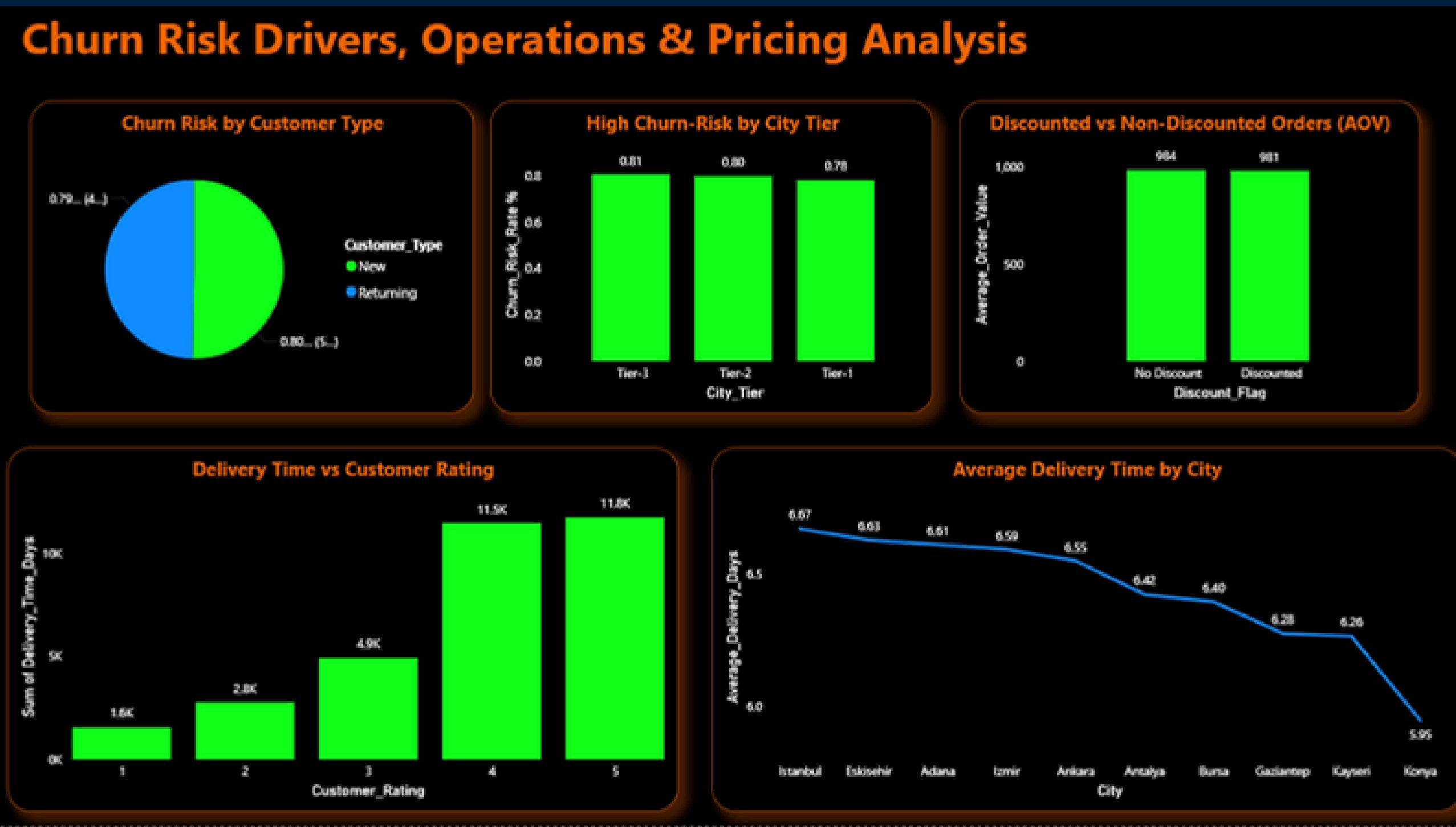
PowerBi Dashboard

Page 2



PowerBi Dashboard

Page 3



Insights From Dashboard

~80% of customers fall under high churn risk, indicating retention is a major business challenge

Returning customers contribute a higher share of revenue compared to new customers, despite being fewer in number.

Revenue shows fluctuations but no strong upward growth trend, suggesting stable but slow business momentum.

Buisness Recommendation

Prioritize customer retention over acquisition by targeting high churn-risk customers with personalized offers, loyalty programs, and re-engagement campaigns, especially among one-time buyers.

Shift marketing and inventory focus toward high-performing cities and product categories (such as top Tier-3 cities and Electronics/Home categories) to maximize revenue impact and operational efficiency.

Re-evaluate blanket discount strategies, as discounts do not significantly improve order value or quantity; instead, use targeted promotions or value-based incentives to protect margins while improving customer experience.

CONCLUSION

This project demonstrates how combining SQL, Python, and Power BI can transform raw e-commerce data into meaningful business insights. The analysis reveals that while customer acquisition is strong, high churn risk and weak retention limit long-term growth. Revenue is largely driven by returning customers, specific cities, and a few high-performing product categories, while discounts and demographic factors such as age have limited impact on spending behavior. Overall, the findings highlight that retention-focused strategies, city- and category-level optimization, and targeted operational improvements offer the greatest opportunity to improve business performance and sustain growth.

Thank
you!