CUSTOMER CHURN ANALYSIS IN E-COMMERCE

Exploratory, Diagnostic & Predictive Analytics using SQL and Python

by

Ademiju Oluwatosin

OBJECTIVES

Identify key drivers of Customer Churn and build a predictive model to proactively flag at-risk Customers.

BUSINESS QUESTIONS

- Do Short-tenure Customers Churn more ?
- Does distance from Warehouse affect Churn ?
- How do complaints influence Churn Risk ?
- Does Satisfaction drop before Churn ?
- Can Churn be predicted in advance ?

Data Overview

- Customer-level e-commerce data
- Target: Churn (0 = Stayed, 1 = Churned)
- Key features: Tenure, Distance, Satisfaction, Complaints & Orders.

TENURE VS CHURN
(SQL INSIGHT)

QUESTION:

Do short-tenure customers churn more or less ?

BUSINESS TAKEAWAY:

Customers with shorter tenure are significantly more likely to churn, indicating early engagement is critical.

```
/* QUESTION 2:
Do short-tenure customers churn more or less? */


SELECT CASE
            WHEN CAST (Tenure AS INTEGER) <= 10 THEN "Short_term"
            WHEN CAST (Tenure AS INTEGER) BETWEEN 11 AND 30 THEN "Mid_term"
            ELSE "Long_term"
            END AS "Cust_tenure",
ROUND(SUM(CASE WHEN Churn = 1 THEN 1 ELSE 0 END) * 1.0 / COUNT(*),2) AS Churn_rate FROM ecommerce_table
GROUP BY Cust_tenure
ORDER BY Churn_rate DESC ;


/* INTERPRETATION OF RESULT 2:
Customer churn follows a non-linear pattern across tenure. Mid-term customers (11-30) exhibit the lowest churn rate, suggesting a stable engagement phase.
Short-term customers show moderate churn, likely reflecting early onboarding risk, while higher churn among long-term customers reflects cumulative exposure
over time rather than immediate dissatisfaction*/
```

	Cust_tenure	Churn_rate
1	Long_term	0.28
2	Short_term	0.25
3	Mid_term	0.05

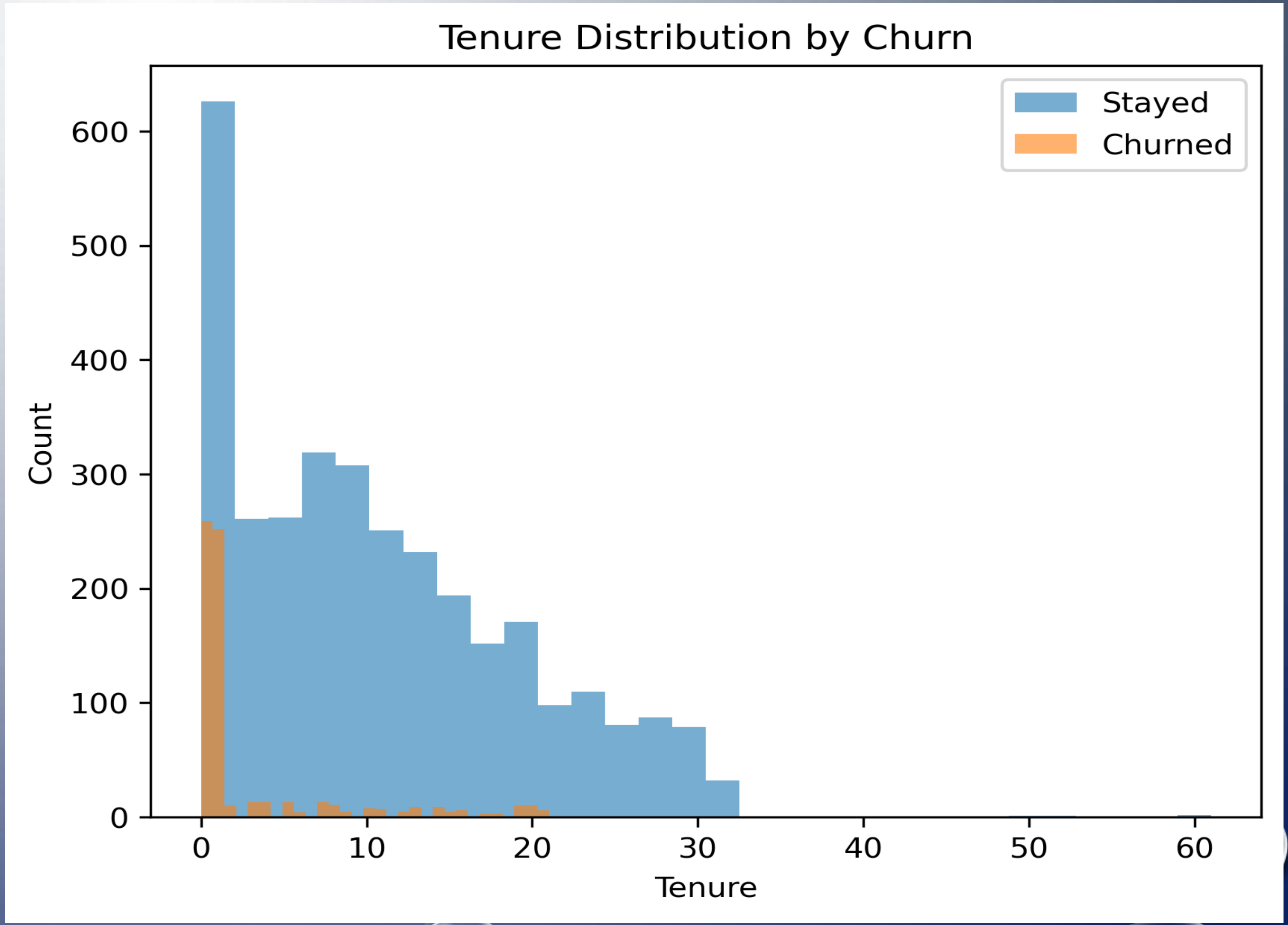
TENURE VS CHURN
(PYTHON CHART)

QUESTION:

Do short-tenure customers churn more or less ?

BUSINESS TAKEAWAY:

The tenure distribution for churned and non-churned customers show a clear right-skewed pattern for both groups. However, churned customers are heavily concentrated at lower values, while non-churned customers exhibit a broader distribution with a longer right tail. This indicates that churn primarily occurs early in the customer lifecycle, and customers who remain active beyond the initial period are significantly more likely to stay long-term. Tenure therefore appears to be a strong predictor of churn risk.



WAREHOUSE DISTANCE CHURN

(SQL INSIGHT)

QUESTION:

Does distance from Warehouse affect Churn ?

BUSINESS TAKEAWAY:

Most customers live close to warehouses, but churn is slightly higher among customers farther away, suggesting delivery experience may influence retention.

```sql
/* QUESTION 9:
Do customers who live farther from the warehouse churn more?*/


SELECT CASE
            WHEN WarehouseToHome <= 40 THEN "Short_distance"
            WHEN WarehouseToHome BETWEEN 41 AND 90 THEN "Mid_distance"
            ELSE "Long_distance" END AS "Distance_to_Warehouse",
COUNT (*) AS Customers,
ROUND(SUM(CASE WHEN Churn = 1 THEN 1 ELSE 0 END),2) AS Churned_customers,
ROUND(SUM(CASE WHEN Churn = 1 THEN 1 ELSE 0 END) * 1.0 / COUNT(*),2) AS Churn_rate

 from ecommerce_table
 GROUP BY Distance_to_Warehouse
 ORDER BY Churn_rate DESC;


/*INTERPRETATION OF RESULT:
Customers who live farther from the warehouse are more likely to churn. This suggests logistics distance and dilivery experience are key drivers of churn.*/
```

	Distance_to_Warehouse	Customers	Churned_customers	Churn_rate
1	Long_distance	170	60.0	0.35
2	Short_distance	3771	614.0	0.16

WAREHOUSE DISTANCE CHURN
(PYTHON CHART)

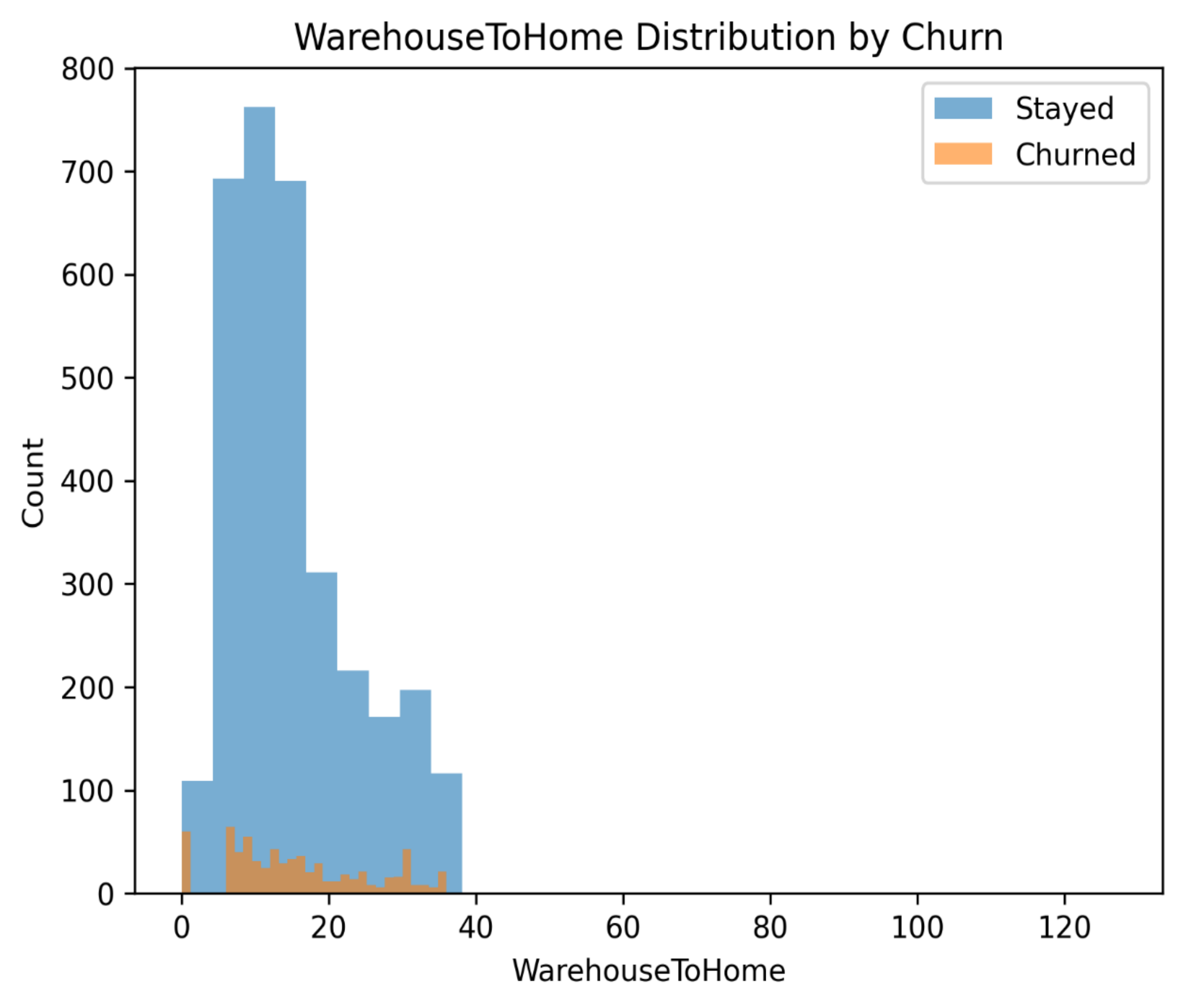
QUESTION:

Does distance from Warehouse affect Churn ?

BUSINESS TAKEAWAY:

Warehouse-to-home distance is not a dominant churn driver on its own. Customers who live closer to the warehouse are far more likely to stay.

Most customers are clustered very close to the warehouse (right skew). There is a long right tail (some customers at very large distances, e.g., 60-120), this means the typical customer lives close, but a small minority lives very far away. Even though logistics distance matters operationally, customers churn regardless of proximity, suggesting that other factors (e.g., tenure, satisfaction score, complaints, engagement) likely play a larger role in churn decisions.



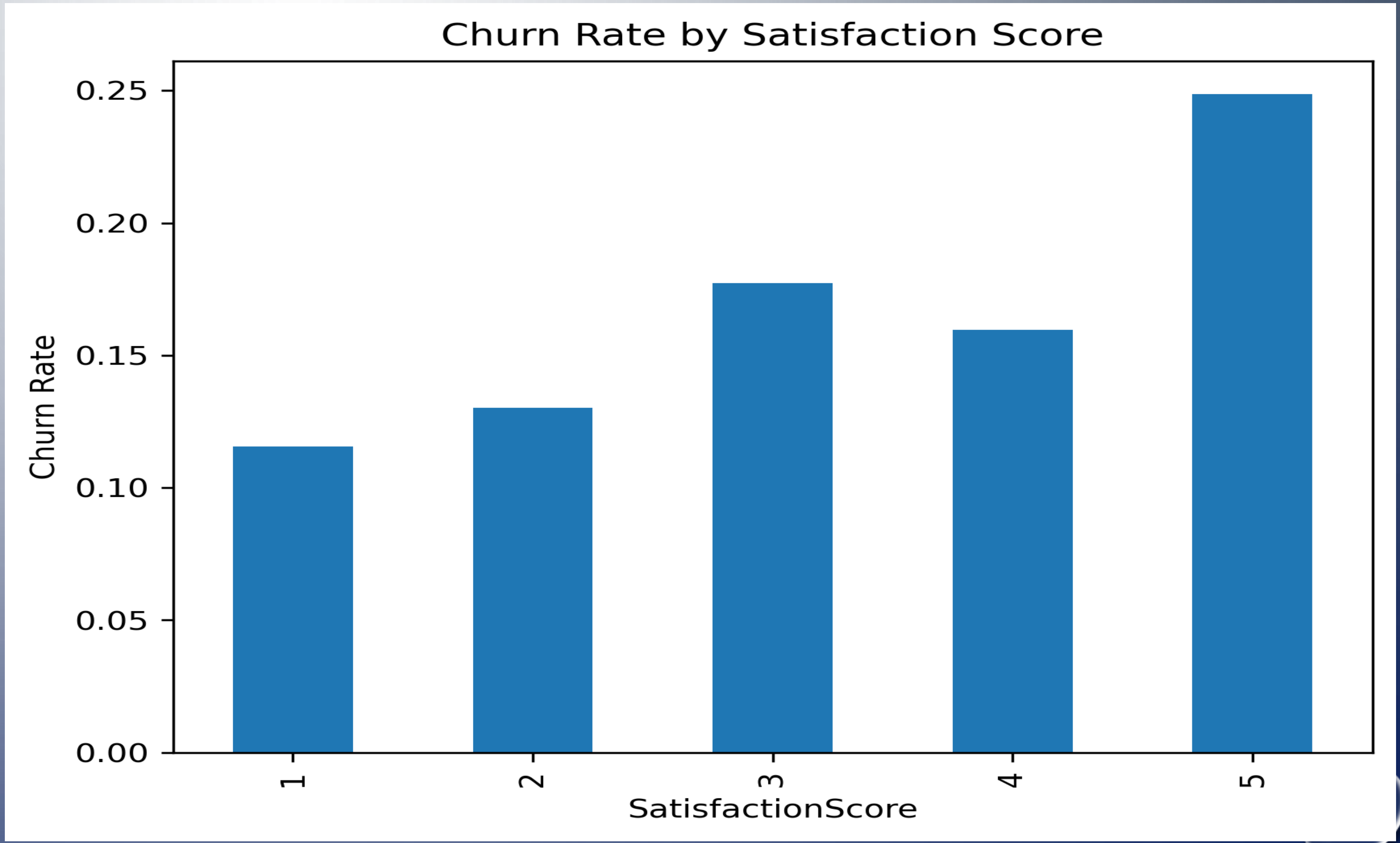
SATISFACTION SCORE VS CHURN
(PYTHON CHART)

QUESTION:

Does satisfaction score drop before they churn?

BUSINESS TAKEAWAY:

The Churn rate increases steadily as the satisfaction score worsens. Customers with the lowest satisfaction (score 5) exhibit the highest churn rate (~24%), while highly satisfied customers (score = 1) have the lowest churn rate (~11%). This indicates that customer dissatisfaction is a strong driver of churn and confirms satisfaction score as a key predictive feature for churn modeling.



COMPLAINT VS CHURN

(SQL INSIGHT)

QUESTION:

Do long-term customers complain more?

BUSINESS TAKEAWAY:

Complaint rates do not exhibit a linear relationship with customer tenure.

```
SELECT
CAST(Tenure AS INTEGER) AS No_of_days, COUNT(*) AS Customers, ROUND(SUM(CASE WHEN Complain = 1 THEN 1 ELSE 0 END) * 1.0 / COUNT (*),2) AS Complain_rate FROM ecommerce_table
WHERE CAST(Tenure AS INTEGER)
GROUP BY CAST(Tenure AS INTEGER)
ORDER BY Complain_rate DESC;


/*INTERPRETATION OF RESULT:
Complaint rates do not exhibit a linear relationship with customer tenure. While extreme rates appear at a very high tenure levels, these are driven by low customer counts
rather than systematic behaviour.  This sugests complaints are triggered by service-related events rather than customer lifecycle stage. */
```

	No_of_days	Customers	Complain_rate
1	61	1	1.0
2	60	1	1.0
3	31	32	0.47
4	29	33	0.45
5	30	46	0.41

COMPLAINT VS CHURN
(SQL INSIGHT)

QUESTION:

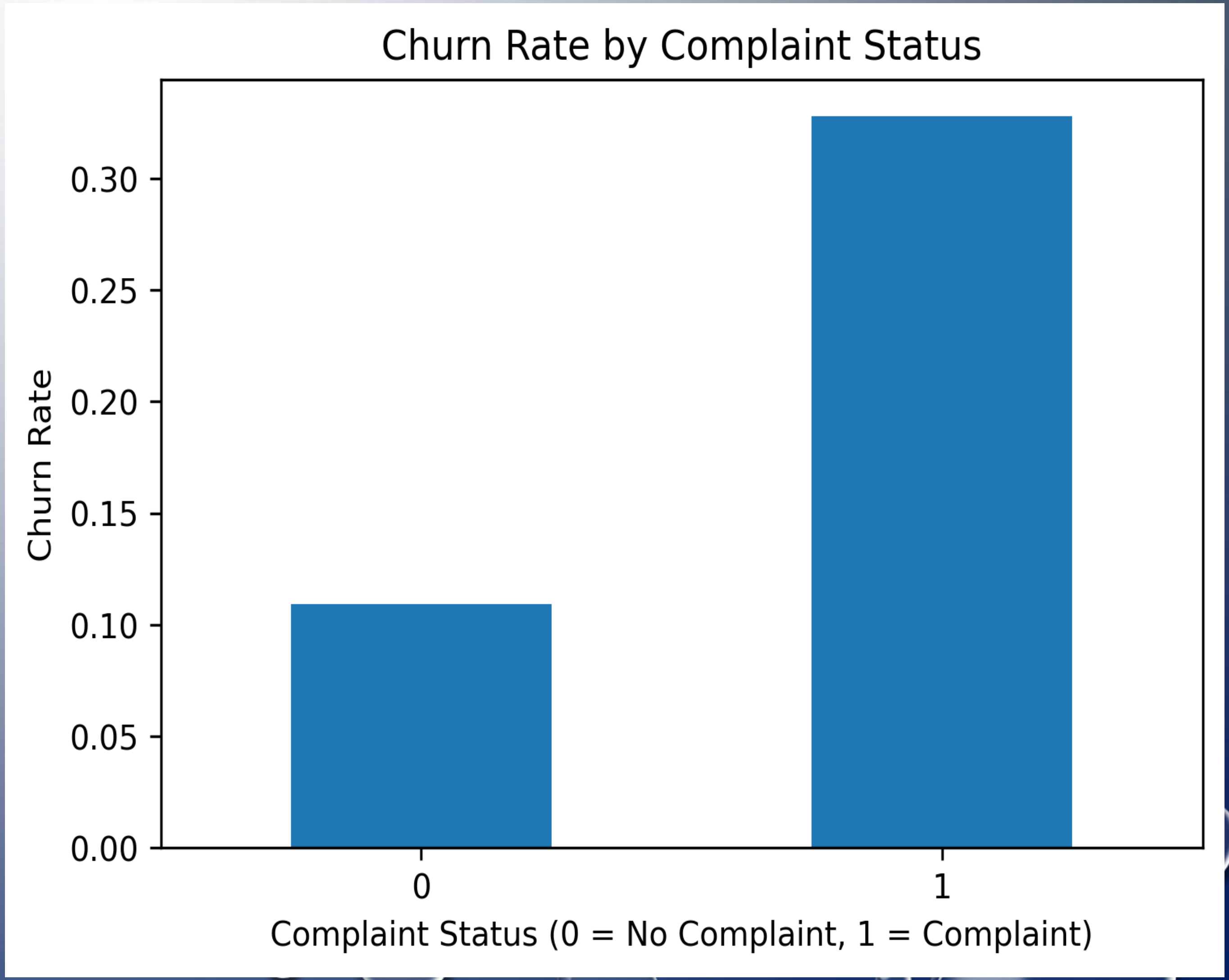
Do long-term customers complain more?

This chart shows a clear relationship between customer complaints and churn behavior.

- Customers who lodged complaints have a significantly higher churn rate compared to those who did not complain.
- - Customers without complaints churn about 11% of the time.
- - Customers with complaints churn about 33% of the time.
- - Customers who complain are about 3x more likely to churn.

BUSINESS TAKEAWAY:

Complaints are a strong early warning signal for customer churn



PREDICTIVE ANALYSIS

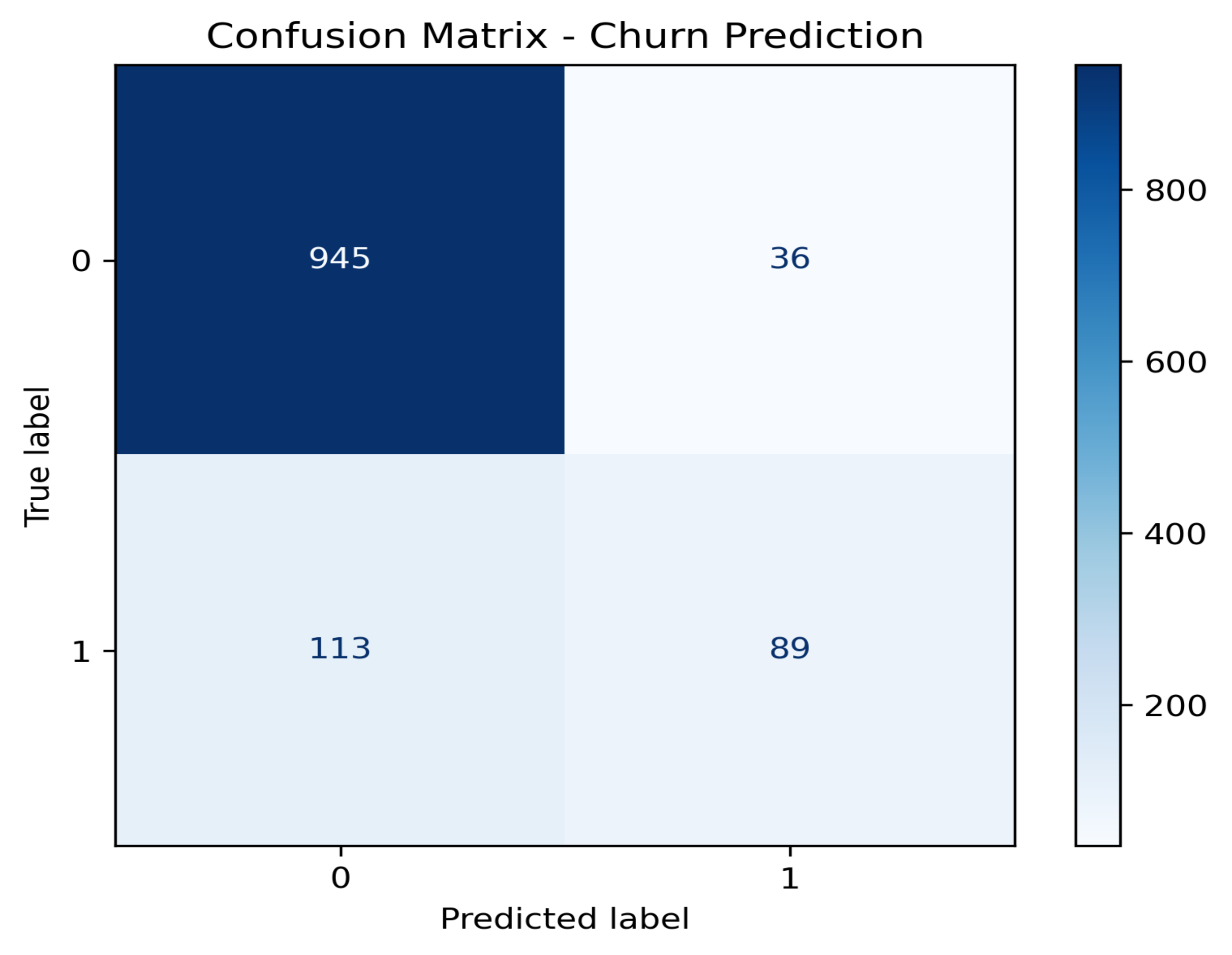
PREDICTIVE MODEL

- Logistic regression : In churn analysis, logistic regression helps us estimate the likelihood that a customer will leave an e-commerce platform based on their past behavior and past behavior and interactions.
- Inputs: tenure, distance, satisfaction, complaints, etc.

MODEL PERFORMANCE
(PYTHON CHART)

- **Confusion matrix**

The model correctly identifies most customers who stay and a meaningful portion of customers who churn, making it useful for targeted retention strategies.



- True Negatives (945): These Are Customers The Model Predicted Would Not Churn, And They Indeed Did Not Churn.
- False Positives (36): These Are Customers The Model Predicted Would Churn, But They Actually Did Not Churn
- False Negatives (113): These Are Customers The Model Predicted Would Not Churn, But They Actually Did Churn.
- True Positives (89): These Are Customers The Model Predicted Would Churn, And They Indeed Did Churn.

CHURN RISK RANKING
(PYTHON CHART)

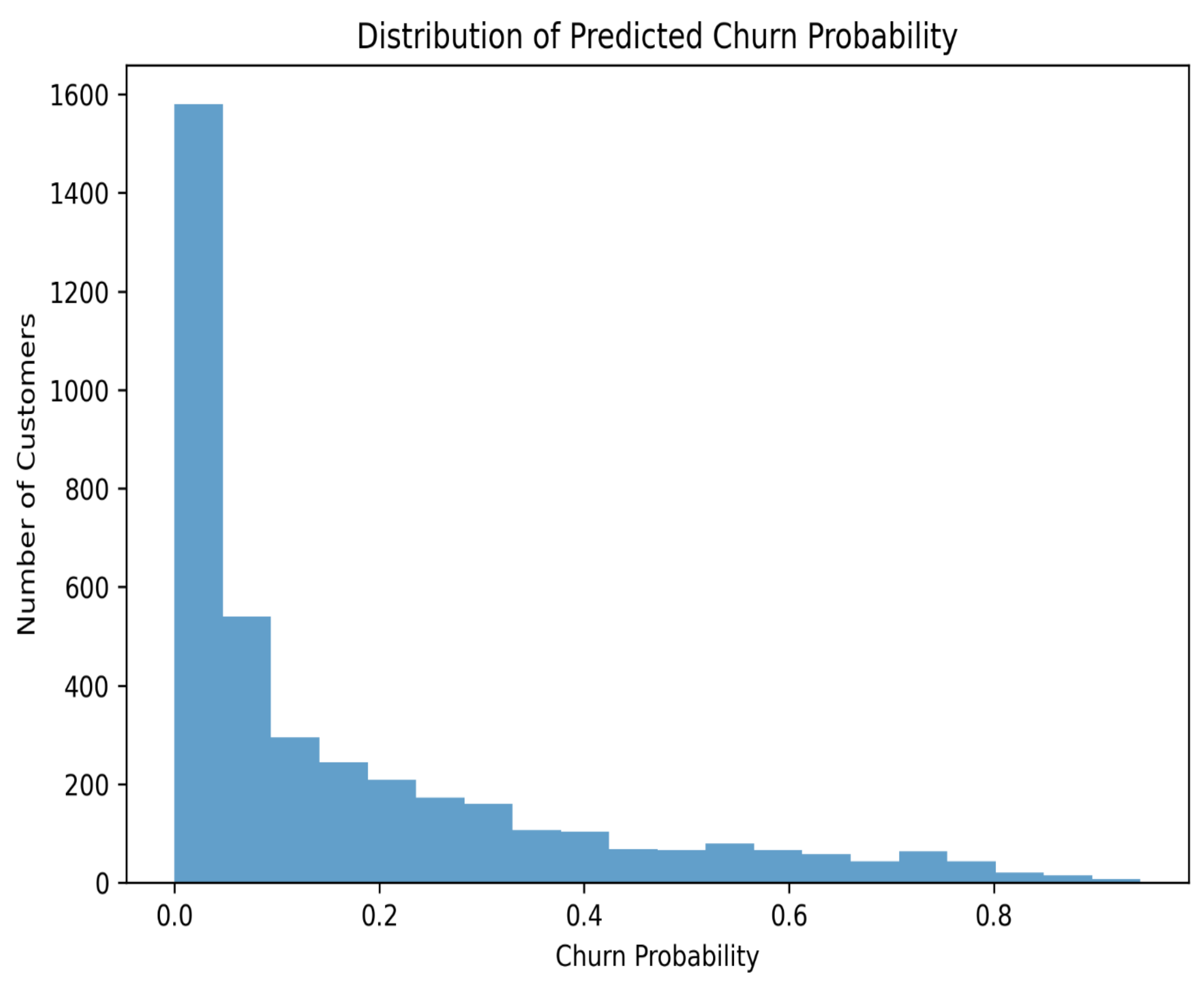
Customers can be ranked by churn risk, enabling proactive interventions for high-risk individuals.

Business Interpretation:

- This Chart shows the distribution of predicted churn probabilities across all customers. Most customers have a low probability of churn, indicating a generally stable customer base. However, there is a noticeable group of customers with high churn probabilities, representing a high-risk segment that requires immediate attention.

Key Insights:

- Churn risk is not evenly distributed.
- Model is useful for prioritization.



RECOMMENDATIONS

1. Focus retention efforts on early-tenure customers.
2. Prioritize fast resolution of customers.
3. Proactively engage customers with high churn probability.
4. Improve experience for customers far from warehouses.

TOOLS USED

- SQL (SQLite / DB Browser)
- Python (Pandas, Matplotlib, Scikit-learn)
- Logistic Regression
- Customer Analytics

THANK YOU