



PURCHASE BEHAVIOR AND CHURN ANALYSIS

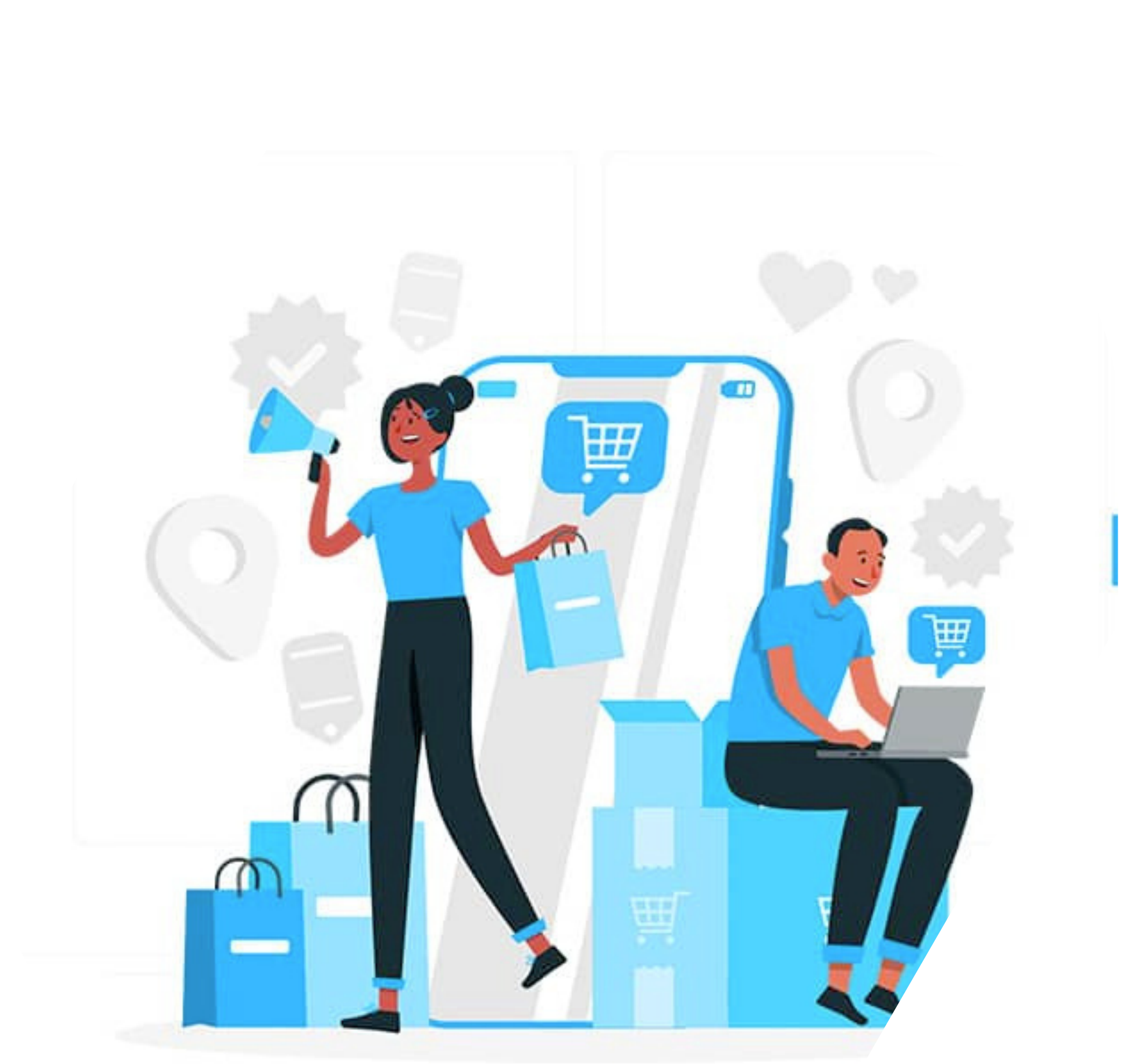
EDA and modeling using transaction data of customers

PRESENTED BY

MICHAEL ANGELO ARSOLON

DATA SOURCE

- Provided by the hiring manager



PROBLEM STATEMENT

Distinguishing businesses using a limited dimension of data (i.e. 3 initial variables) and defining and creating a churn predicting model

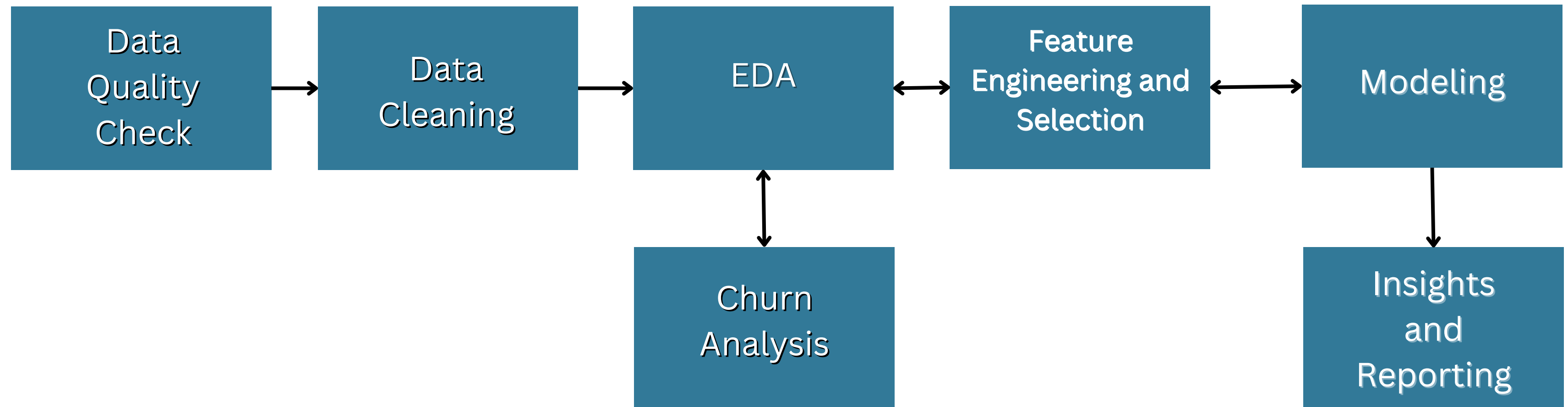
Scope of the study

The scope of this study is to analyze a dataset containing transaction activities of future merchants using a company over a 2-year period (2033-2034). The study aims to identify different types of businesses represented in the sample, define and identify churn, and build a model to predict which active merchants are likely to churn in the near future.

Relevance of the study

The study is relevant to businesses that process transactions and want to identify different types of businesses that use their services, as well as predicting which active merchants are likely to churn. Understanding the patterns of merchants' payments activities can help businesses tailor their products and services to specific merchant types and improve their retention strategies.

PIPELINE



SOME QUESTIONS FOR EDA

Who are the most active merchants based on number of transactions?

Are there particular days of the week when sales peak?

Are there particular month of the year when sales peak?

Can we find high value customers (high price per transaction)?

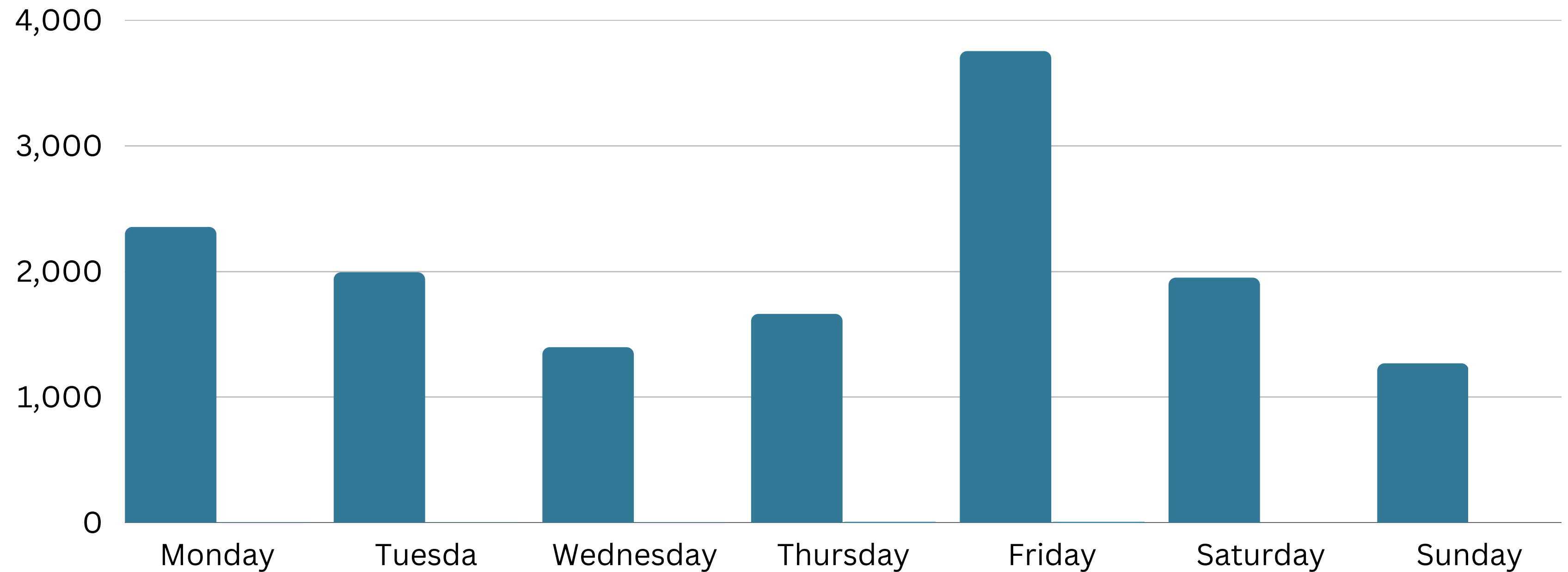
Are there correlations between features generated?

EDA

**Focused on creating merchant profiles based on the data
and do the exploration from there.**

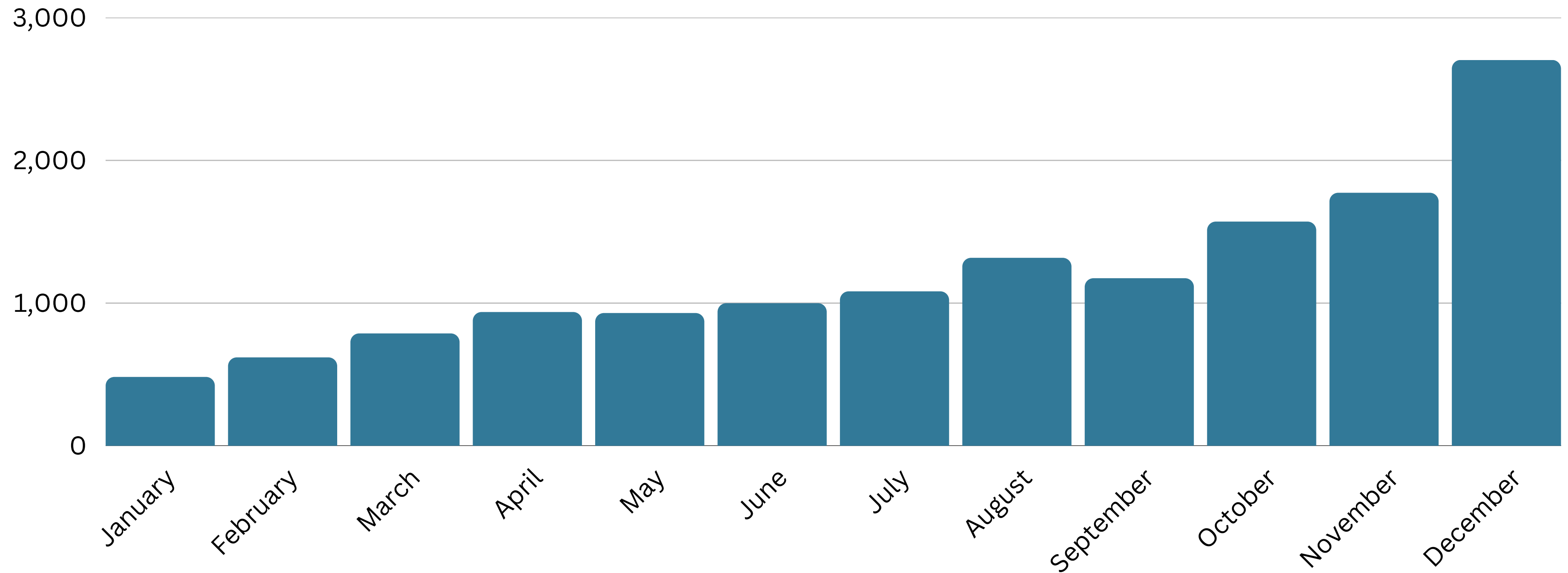
PEAK DAYS

Sales peak during Fridays as expected. These are the day a unique merchant peaks the first time.



PEAK MONTHS

Sales peak during December as expected. These are the month of the years unique merchant peaks the first time.



CHURN ANALYSIS

How was churn defined

1. Get the last and second-to-the last purchase date of each merchant if they have more than 1 transaction.
2. Get the time gap between them
3. The churn is the average of all the gaps

Churn definition

If a customer did not buy after 35 days or 1 month

54%
CHURN

**FEATURES
SELECTED**

**Total number of
transaction**

**Total Price the merchant
spent**

Value: Price per transaction

**Day of the week with
highest number of
transaction**

**Month of the year with
highest number of
transaction**





Modeling

LOGISTIC REGRESSION

With default hyperparameter

	PRECISION	RECALL	F1-SCORE
NOT CHURNED	84%	58%	69%
CHURNED	73%	91%	81%
ACCURACY			76%
SUPPORT 24/7	76%	76%	75%

CONCLUSION AND RECOMMENDATION

Conclusion:

- Churn rate is high at 54%
- Logistic regression model has 76% accuracy and 91% churn recall

Recommendation:

- Improve the current model further
- Explore other churn definitions
- Invest more time in researching and testing other models

The company is experiencing a high churn rate, with 54% of customers leaving the company based on the average gap between their last two sales. This is a cause for concern, as it indicates that the company is losing a significant portion of its customer base.

However, the logistic regression model that has been implemented shows promise, as it has a 76% accuracy rate and a 91% churn recall rate. This means that the model is correctly identifying 91% of the customers who are likely to churn, which is a crucial metric for the company to focus on.

Despite this, I would recommend that the company continues to explore ways to improve the model further. While a 76% accuracy rate is a good start, there may be ways to increase this accuracy even further, which could lead to even better results for the company. Additionally, the company should consider trying other churn definitions to see if they provide more accurate insights into the problem.

Furthermore, I would advise that the company invests more time and resources into researching and testing other models. There may be other machine learning algorithms or statistical techniques that could provide better insights into customer churn, and it is essential to explore all options thoroughly. By doing so, the company can ensure that it is using the best possible model to identify and address customer churn, which can have a significant impact on the company's bottom line.

In summary, the company should continue to work on improving the current model, explore other churn definitions, and invest more resources into researching and testing other models. By doing so, the company can gain a more comprehensive understanding of customer churn and take appropriate action to mitigate it.