Data PORTFOLIO

DUY CAO

project
PORTFOLIO

01
Pizza Hut Analysis
Analyzing and Story-telling
“How to grow market?”

02
Brazilian E-commerce Analysis
Extracting and visualizing
Advanced Tableau - Python

03
Other projects
5+ Projects from Basic
analytics to Deep learning

PIZZA HUT

Link to final product: Report Slides

PROJECT GOAL

Utilize **Python** and **Tableau** to extract **insights** about Pizza Hut's **customers' behaviors** and provide **recommendations** regarding tactics to **retain and grow** their **market.**

SKILLS

- **Data analytics**
- **Data visualization**
- **Presenting**
- **Mathematics**
- **Business Acumen**

KEY TASKS

- Performed **k-means** to classify customer segments.
- Used **Tableau** to extract insights from each segment with visualizations.
- Calculated Customer Lifetime Value (**CLV**).
- Predicted **churn rate** and customer size using an advanced **deep learning** model.
- Revealed 4 factors strongly affecting churn rate using **logistic regression.**
- Spotted **5+ actionable insights** and adopted **storytelling** to present them.

TOOLS

- **Python:** Pandas, Tensorflow,...
- **Tableau**
- **Google Slide**

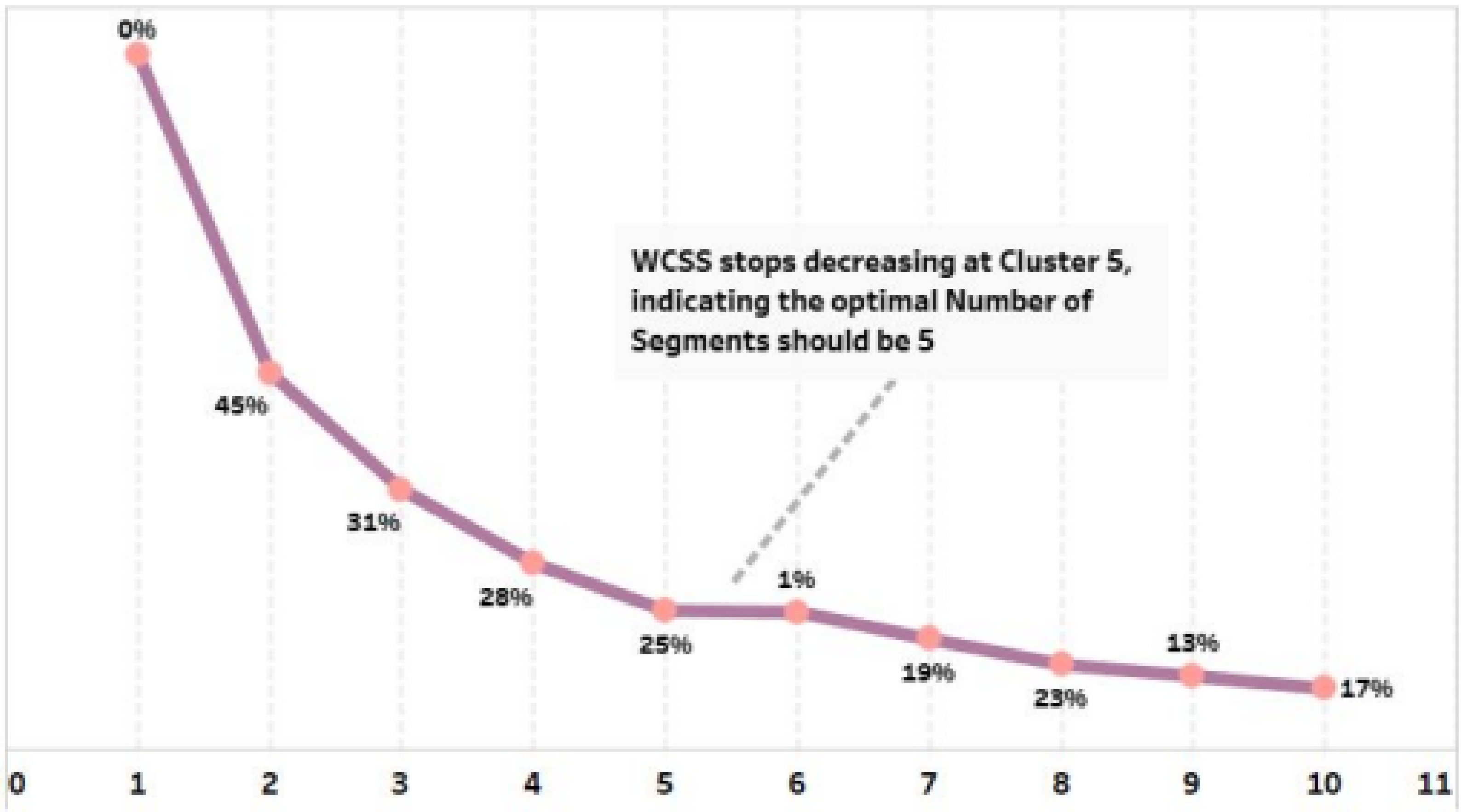
PERFORMED K-MEANS TO CLASIFY CUSTOMER SEGMENT

Clustering from RFM metrics using K-mean method produced 5 patterns of customer behavior

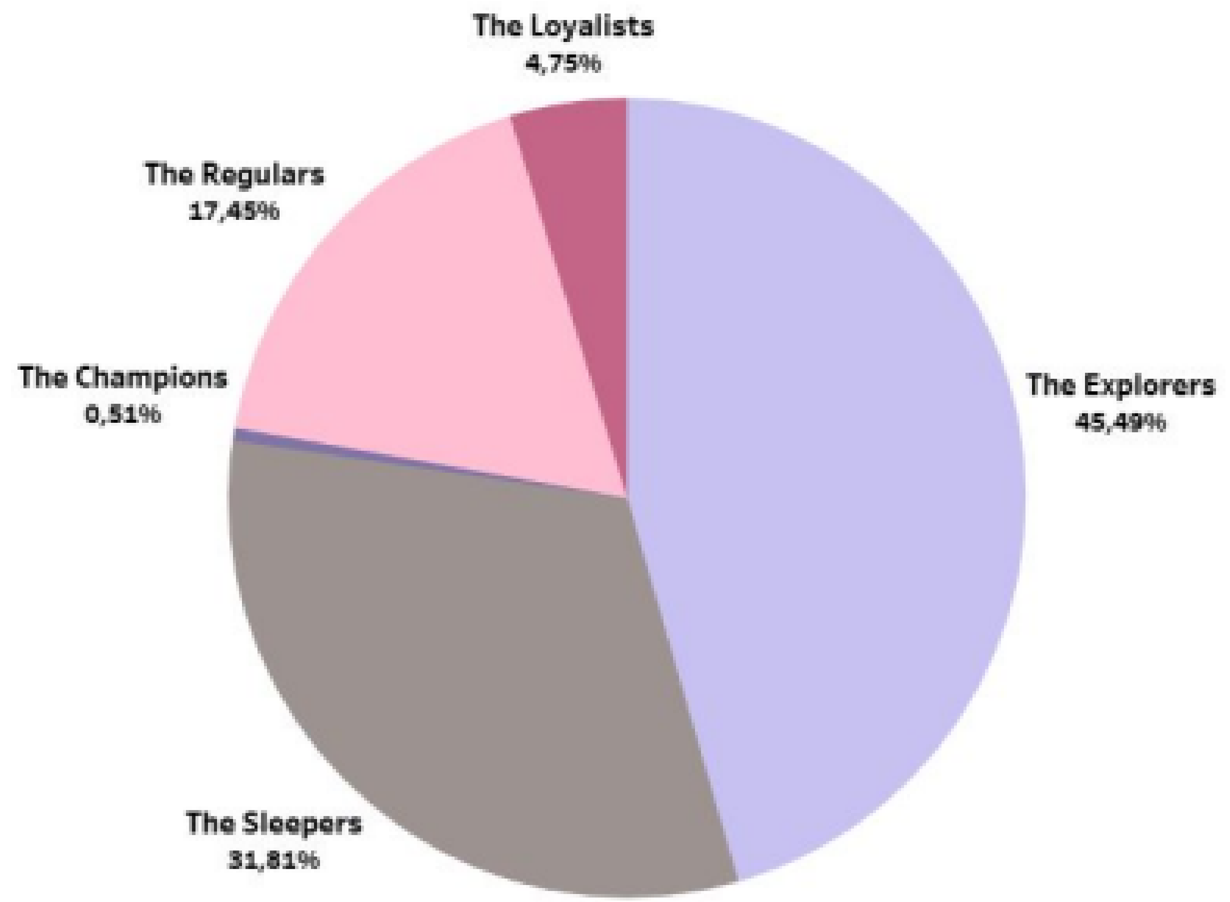
Full Slides Here

RFM metrics	Variable Used	Question to answer
Recency	Days Between Last Order And Current Day	When was the last time they purchase from us?
Frequency	Average Days Between Purchase	How frequently do they use our service?
Monetary	Value Per Order	How much they buy each time?
	Total Value Generated	How much value they have contributed?
LifeTime	LifeTime	How long have they been with us?

Results from K-mean suggesting 5 distinct segments.
Showing *Percentage of WCSS reduced*



5 Groups are generated and named according to their uniqueness.
Showing *Value Contribution of Each Segment*



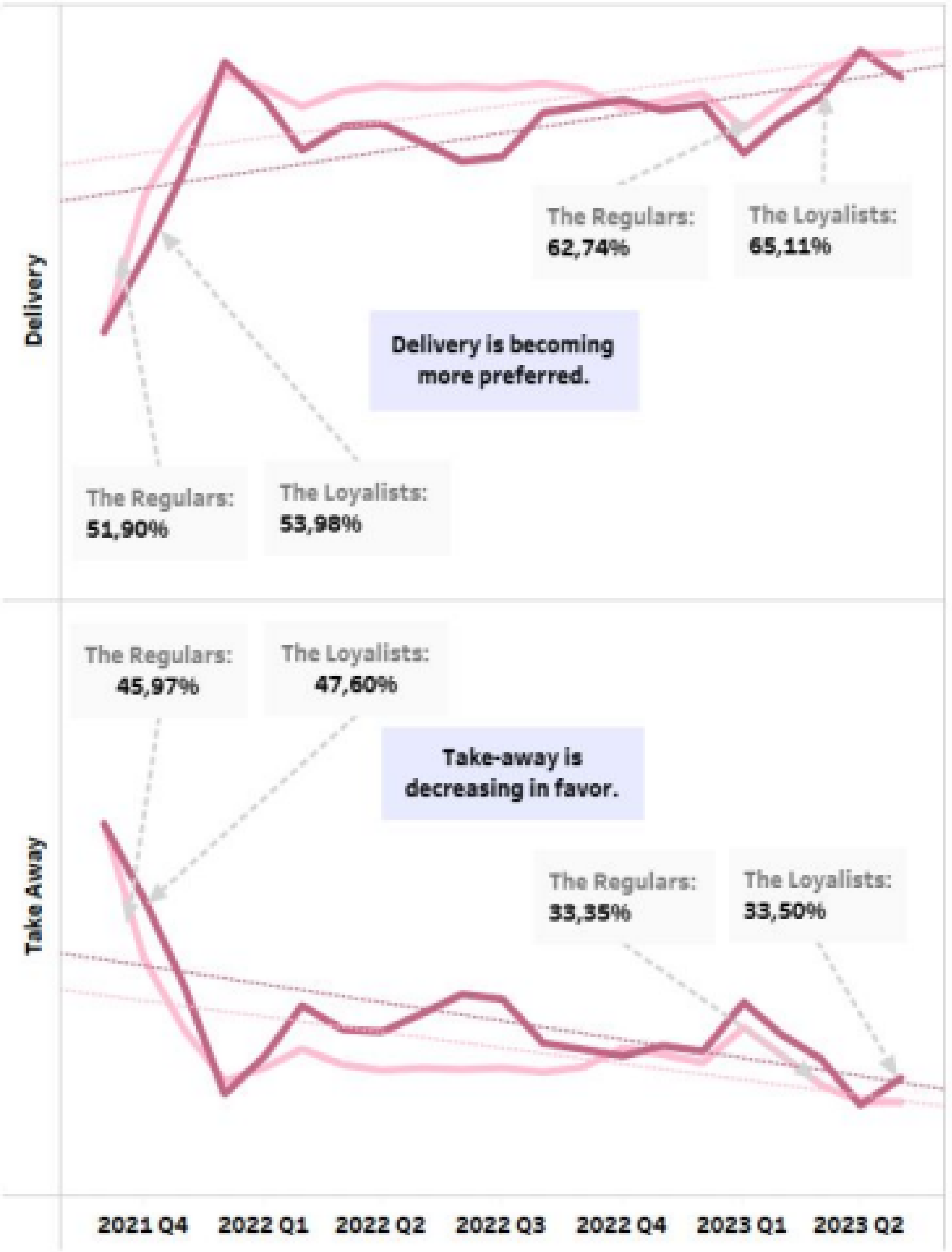
PIZZA HUT

USED TABLEAU TO EXTRACT INSIGHTS FROM EACH SEGMENT

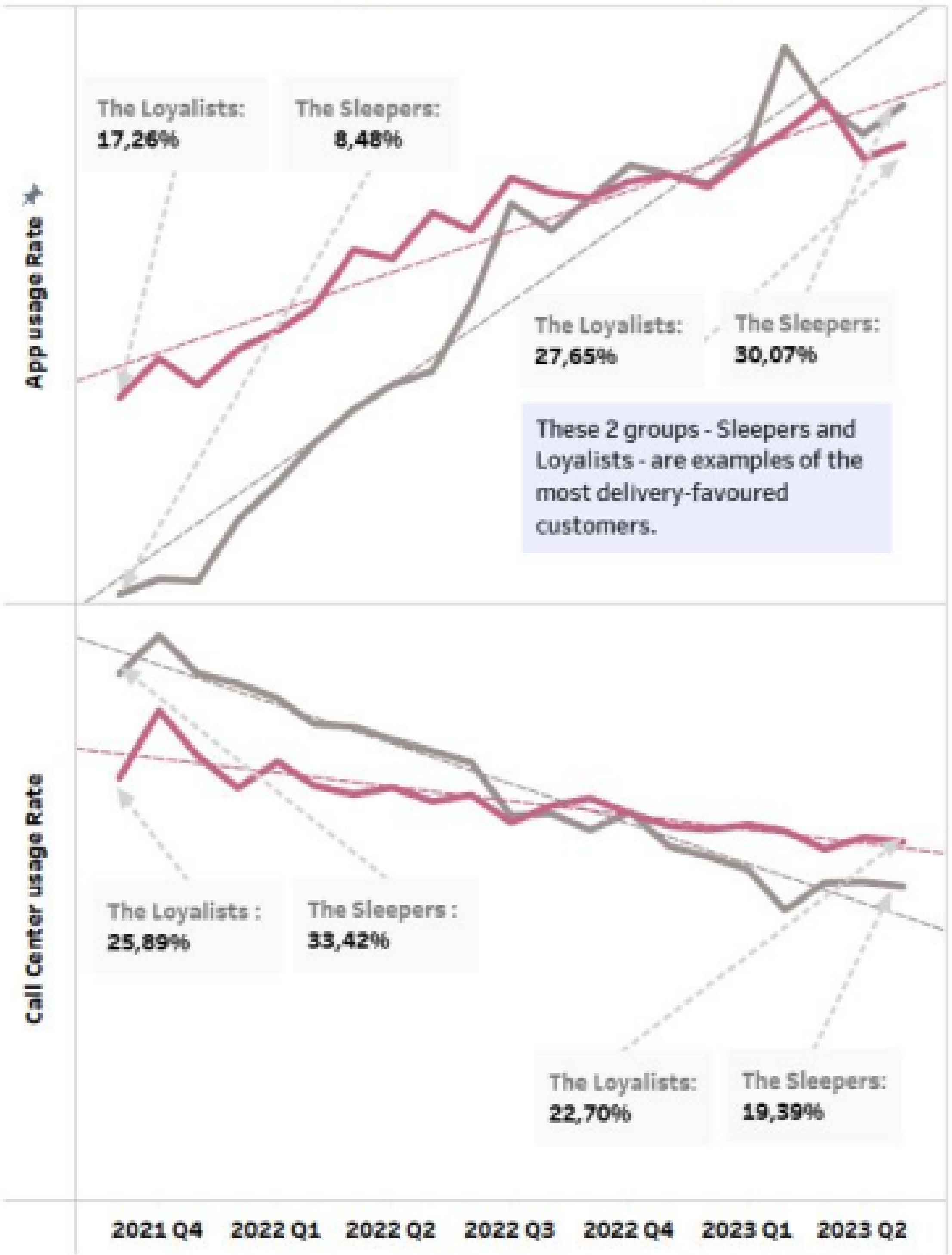
Time series for each segment reveals notable trends in channels, sources & voucher usage.

Full Slides Here

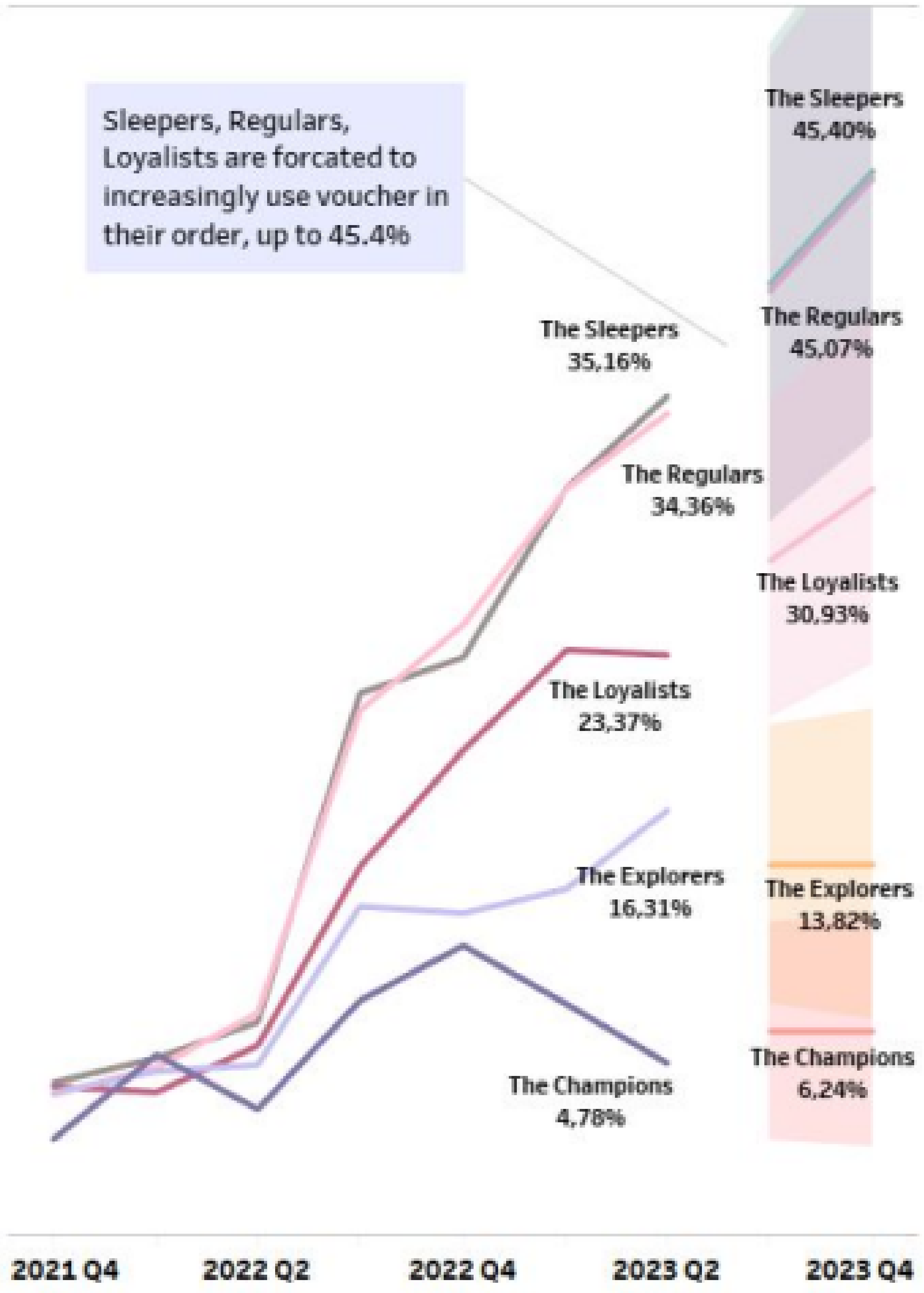
Preference for delivery is on the rise among high-CLV groups like The Loyalists and The Regulars, while the oposite is true for take-away.
Showing *Channel's usage percentage of each segment*



When it's delivery, there is a growing preference for using the app. Meanwhile, delivery through the call center is losing popularity.
Showing *% usage of the app and call center in delivery.*



These Delivery-preferred segments also have an increasing use of vouchers.
Showing *% usage of voucher*



CALCULATED CUSTOMER LIFETIME VALUE (CLV)

Full Slides Here

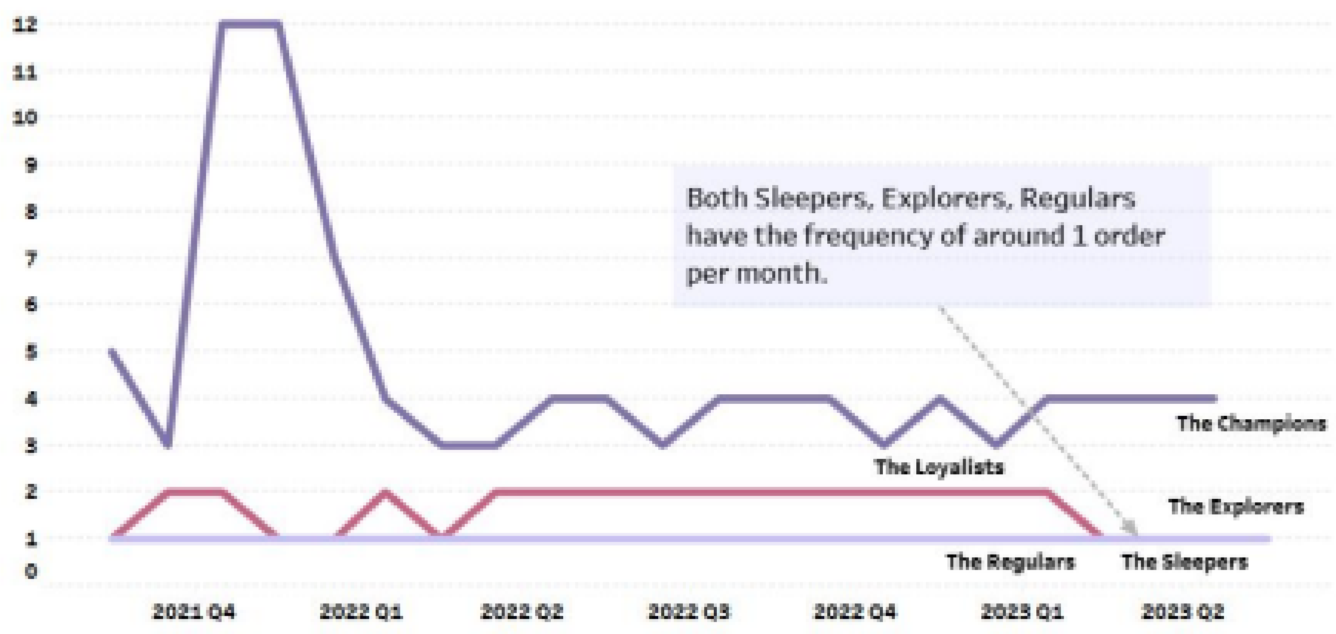
CLV* is calculated based on order frequency, value, and customer lifetime. Outstanding contribution from the Champion segment is evident.

**Customer Lifetime Value (CLV) is the predicted total revenue a business expects to earn throughout its entire relationship with a customer*

**Formula: (Number of orders each month * Value per order) * Expected lifetime in months*

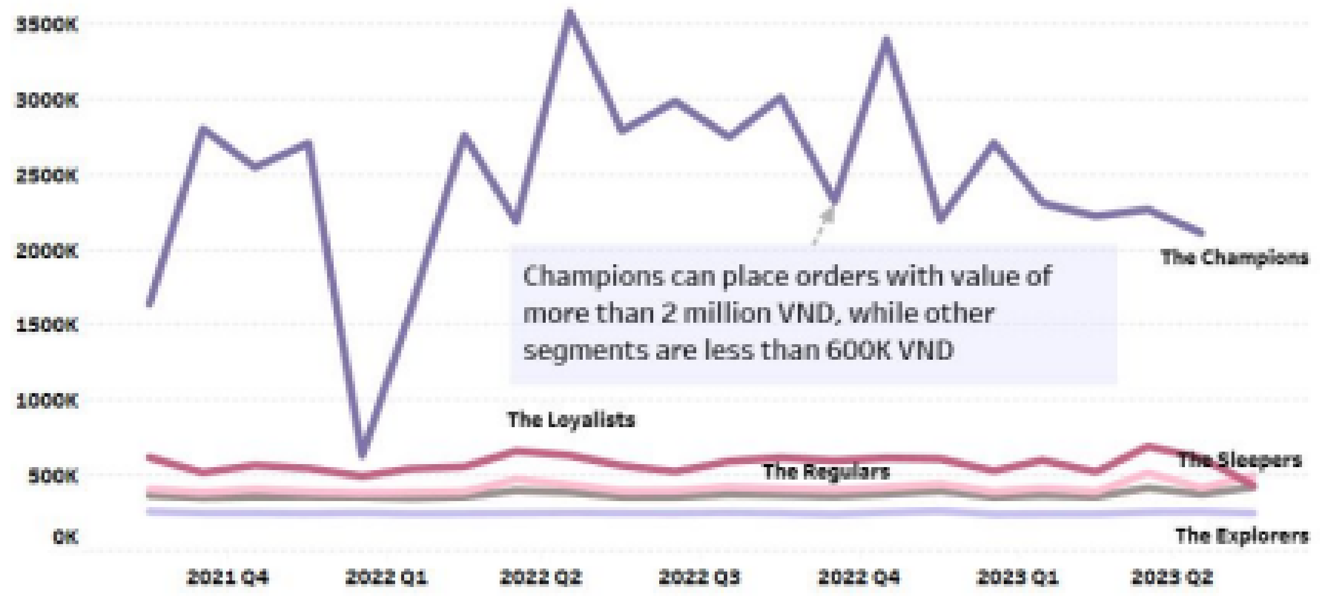
Order Frequency

Number of order placed per month



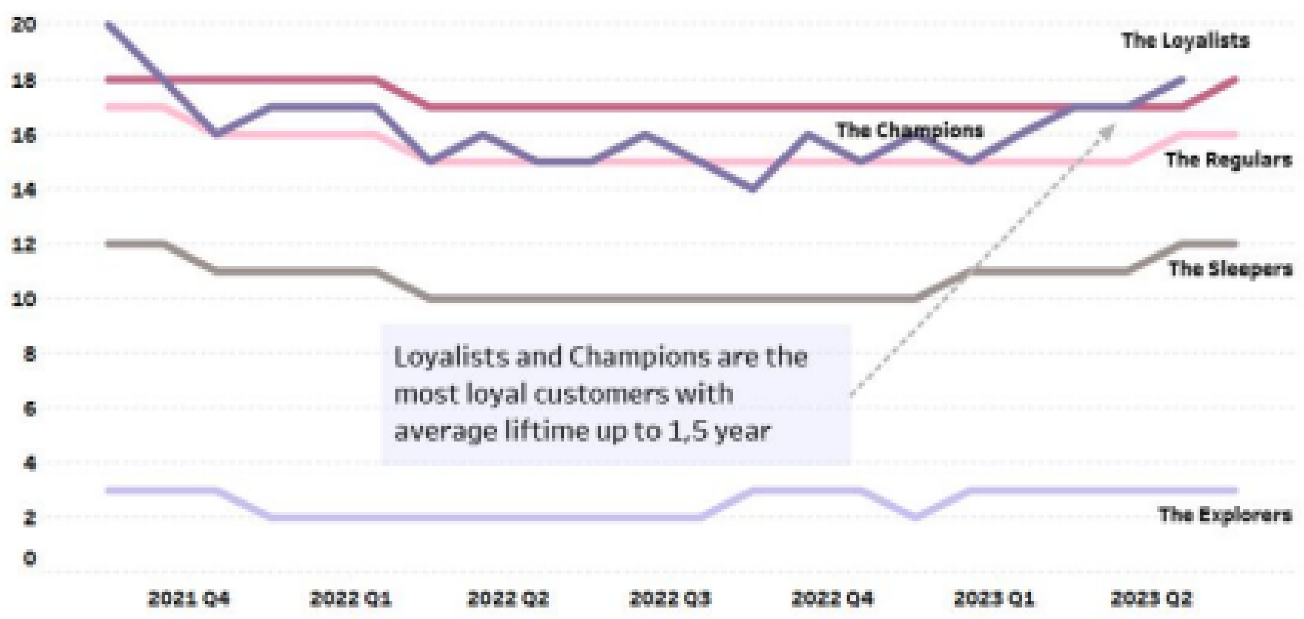
Order value

Value per order



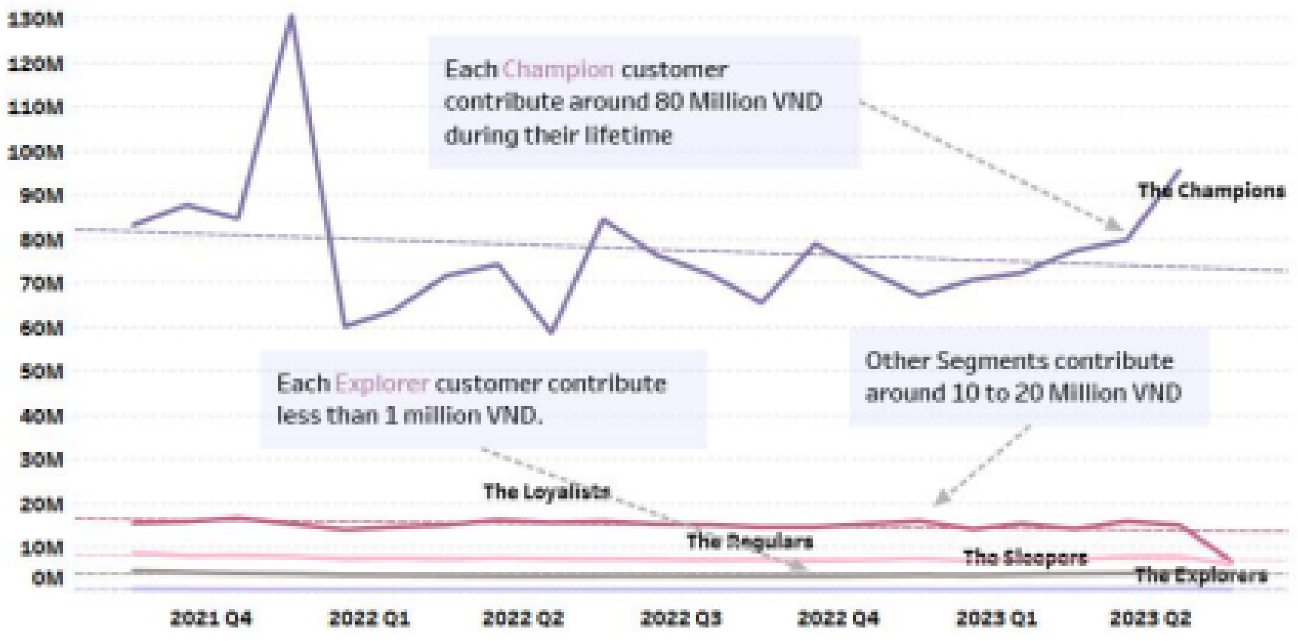
Customer Lifetime

Number of months since first to last order



Customer Lifetime Value

Order frequency * Order value * Customer Lifetime



PREDICTED CHURN RATE AND CUSTOMER SIZE WITH ADVANCED DEEP LEARNING MODEL

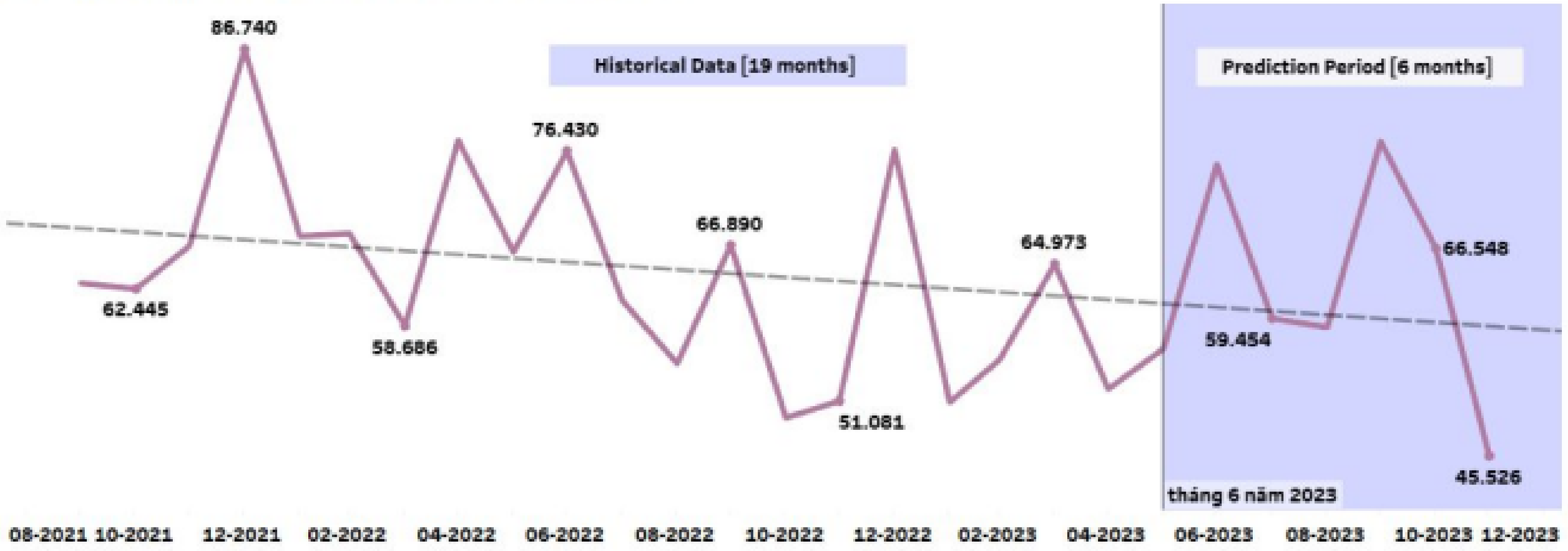
Deep learning models were applied to forecast future customer size and understand churn* patterns, with results showing a decrease in customer numbers.

Churn rate: the percentage of customers who order in one month but disappear the next.

Full Slides Here

Past data and the model's predictions indicate a declining trend in customer size.

Showing *total number of customers at each time*

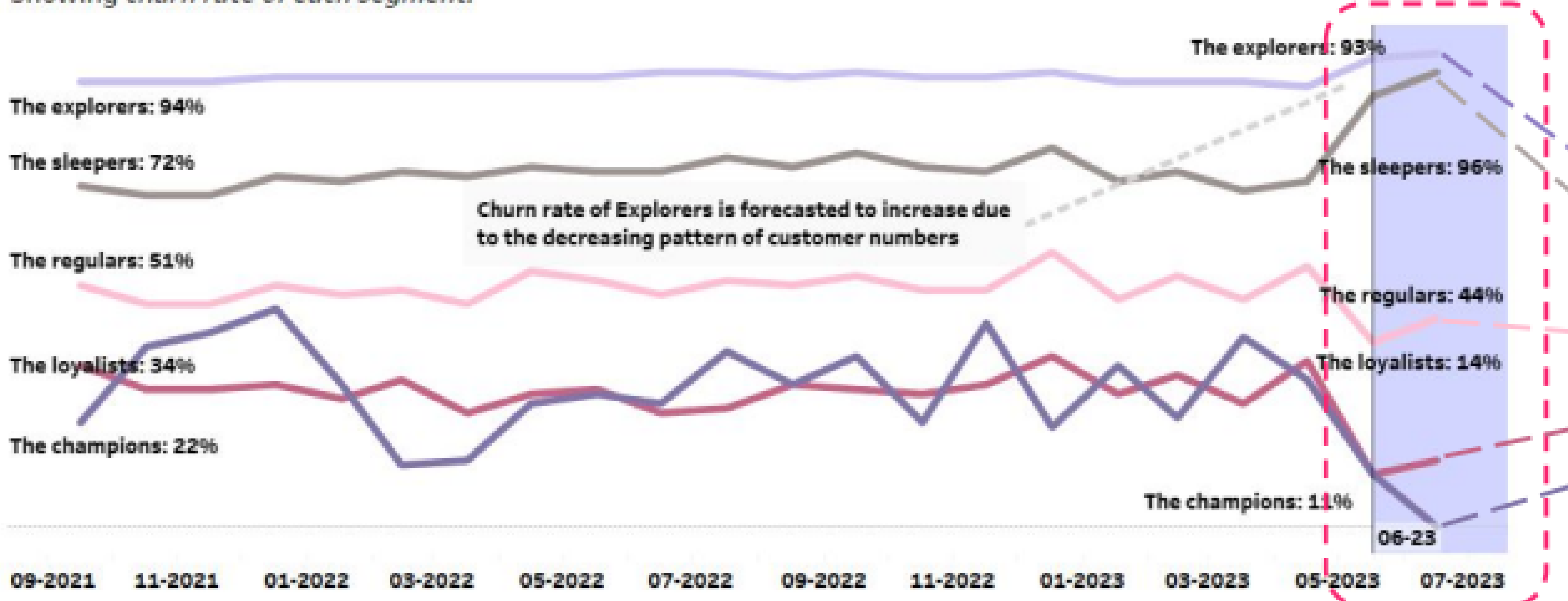


How the prediction was done

Predictive question	Model's input	Model used
How many customers do we have each month? [Customer size]	**Time series data:** total number of active customers each week/month	Bidirectional-LSTM RNN
Will this customer return next month? [Churn rate/ Return rate]	**Every provided information**, 10+ variables: RFM metrics, CLV, purchasing habits, demographic, etc.	Neural network with 10+ dense layers

Most segments exhibit stable churn rates. Return rate* of current customers is also predicted.

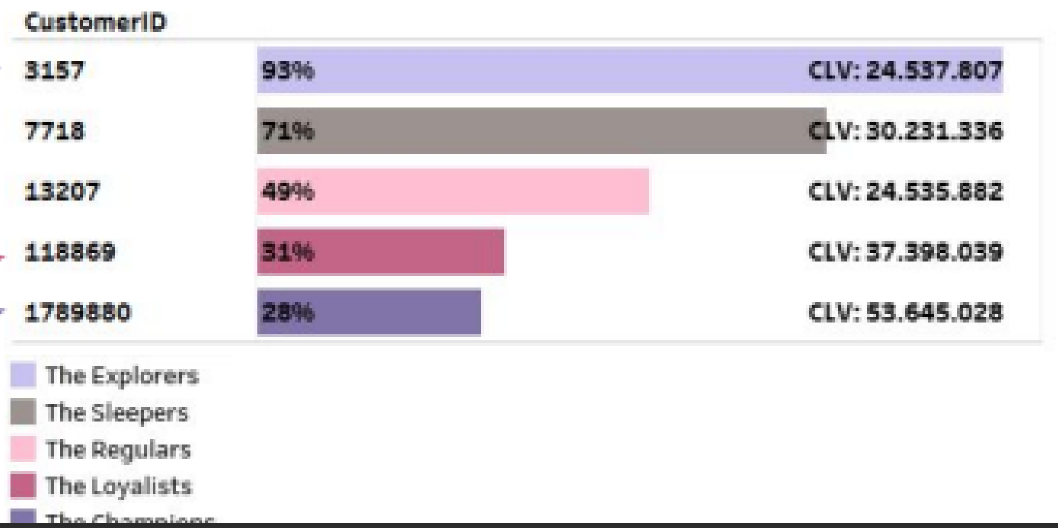
Showing *churn rate of each segment.*



Return rate: the percentage of customers ordering in one month and the next.

Sample customers from each segment and their Churn rate | CLV

Showing *Churn Rate in 07-2023 of Each Segment | CLV*



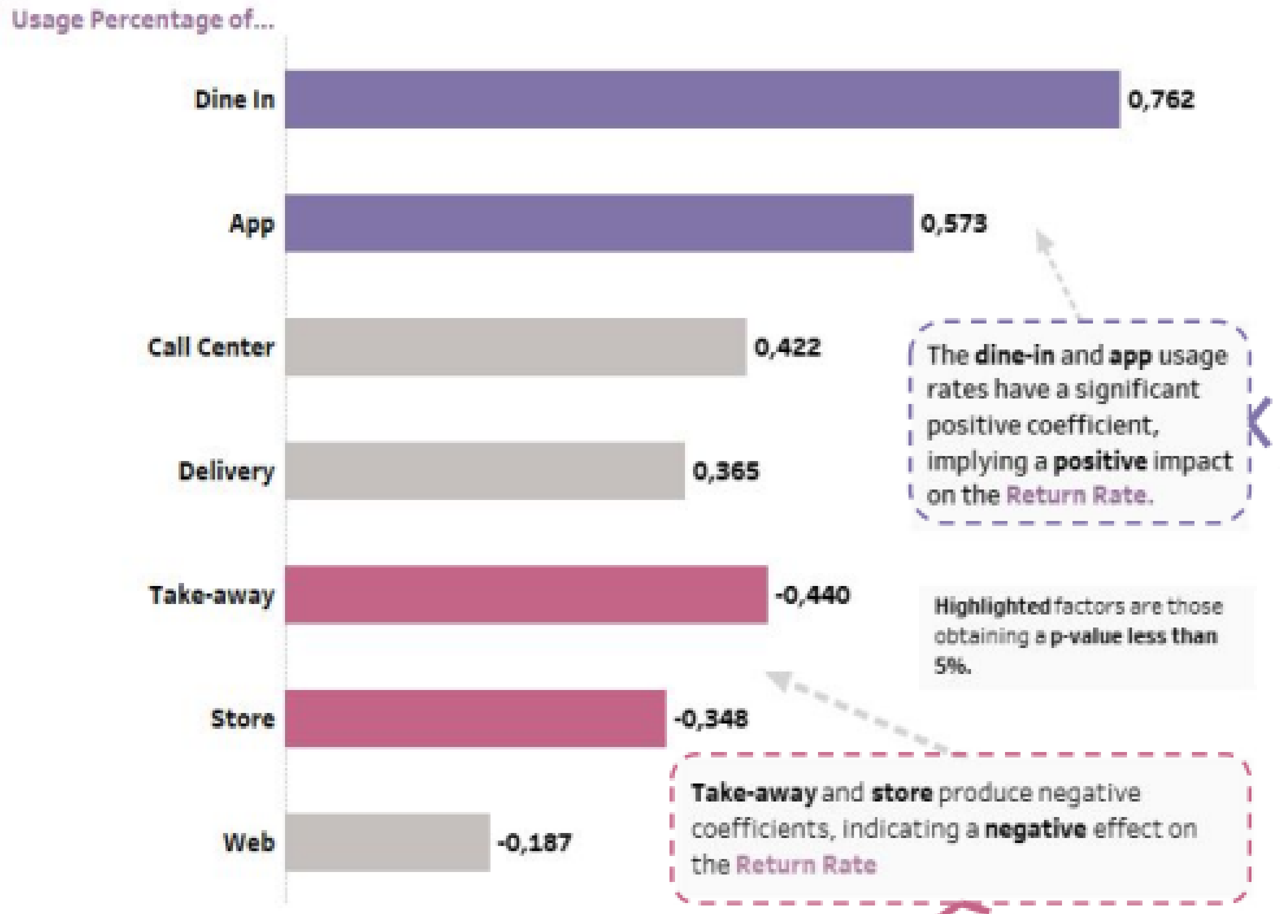
REVEALED 4 FACTORS STRONGLY AFFECT CHURN RATE USING LOGISTIC REGRESSION

Logistic regression reveals a correlation between customer Return Rate* and 4 factors from order channel and source

Full Slides Here

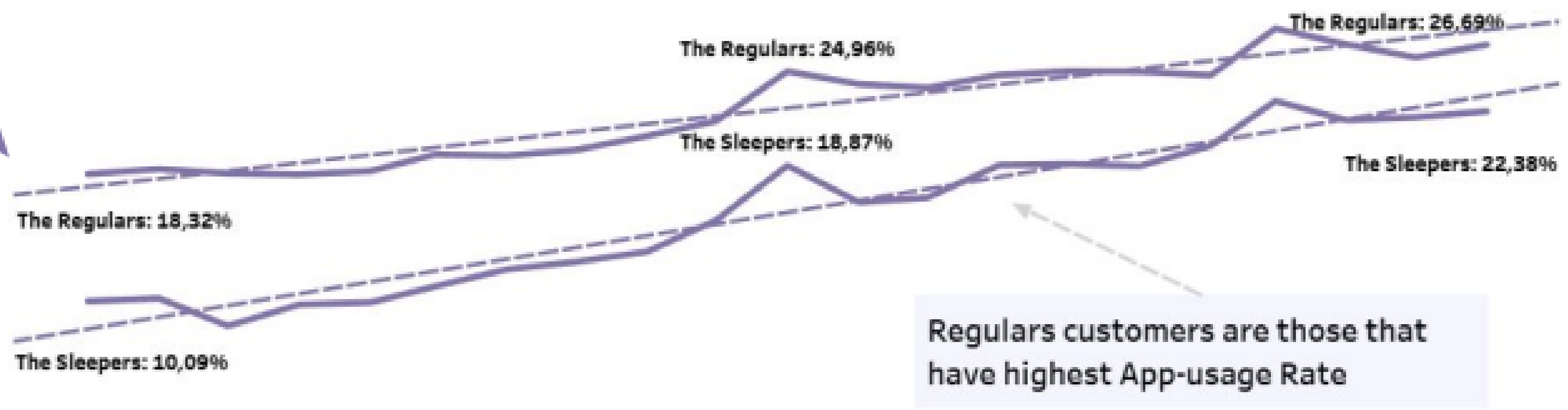
*Return rate: the percentage of customers ordering in one month and the next.

4 variables demonstrating a strong effect on the Return Rate are spotted.
Showing *the coefficient of each variable on Return Rate*



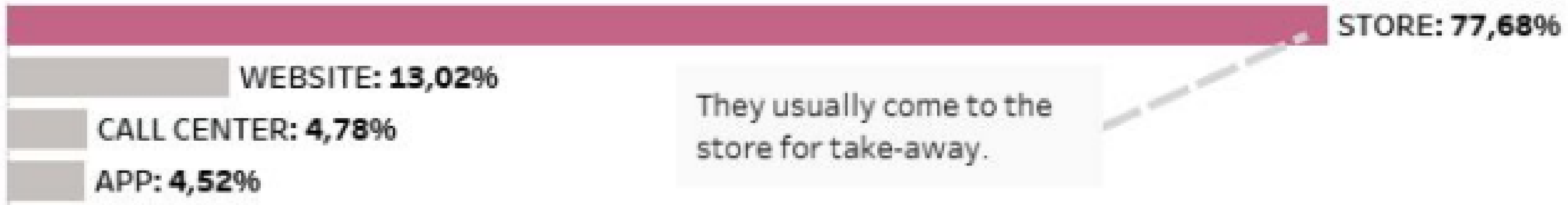
Customers who have **high Dine-In rate** have **higher Return Rate.**
Loyalists and Champions are great examples.
Showing *Dine-In Rate of each Segment*

Higher app usage frequency reduces the likelihood of leaving.
Coupled with increasing app adoption, there is a compelling reason to invest in this platform.
Showing *App Usage Rate of each Segment*



Customers who **prefer Take-away** have **lower Return Rate.**
Explorers are the main driver of this phenomenon.
Showing *Take-away Rate of each segment*

Since they usually come to store when taking away,
customers who have **high 'go-to-store'** rate also have **lower Return Rate**
Showing *'Go-to-store' when take-away rate across all segments.*



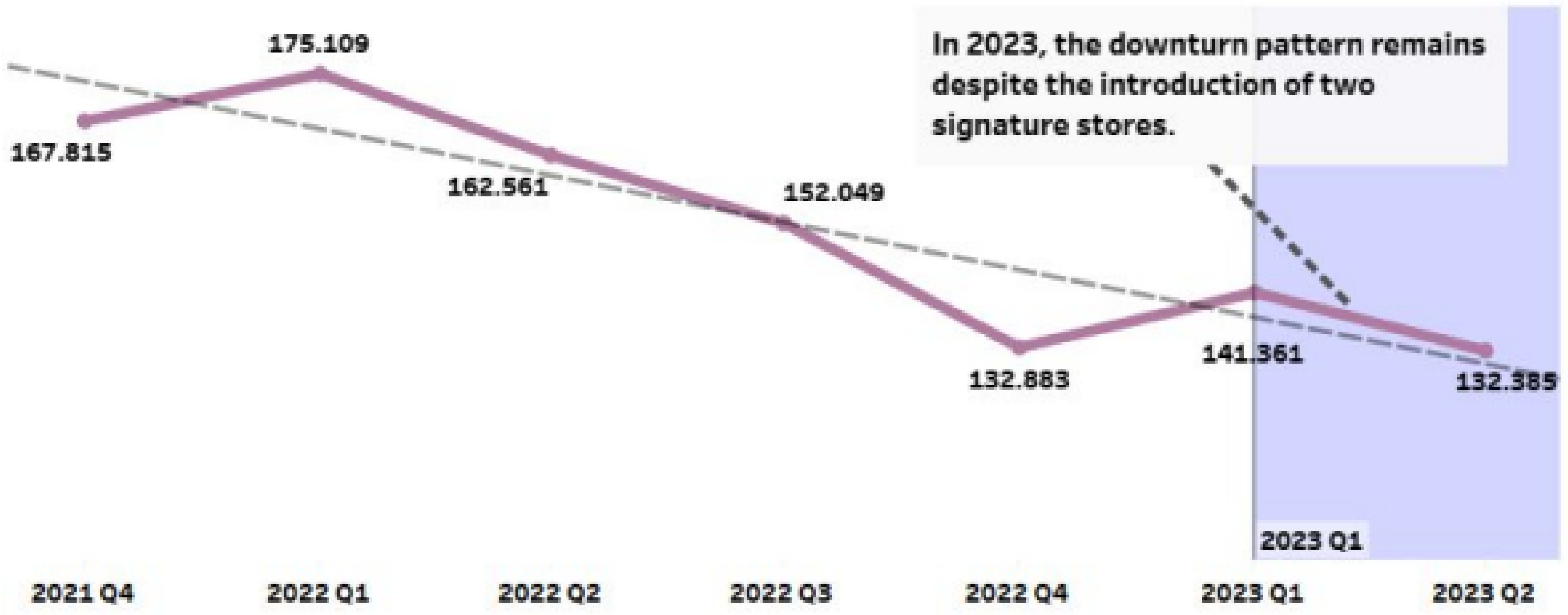
PIZZA HUT

SPOTTED 5+ ACTIONABLE INSIGHTS* AND ADOPTED STORY-TELLING TO PRESENT THEM

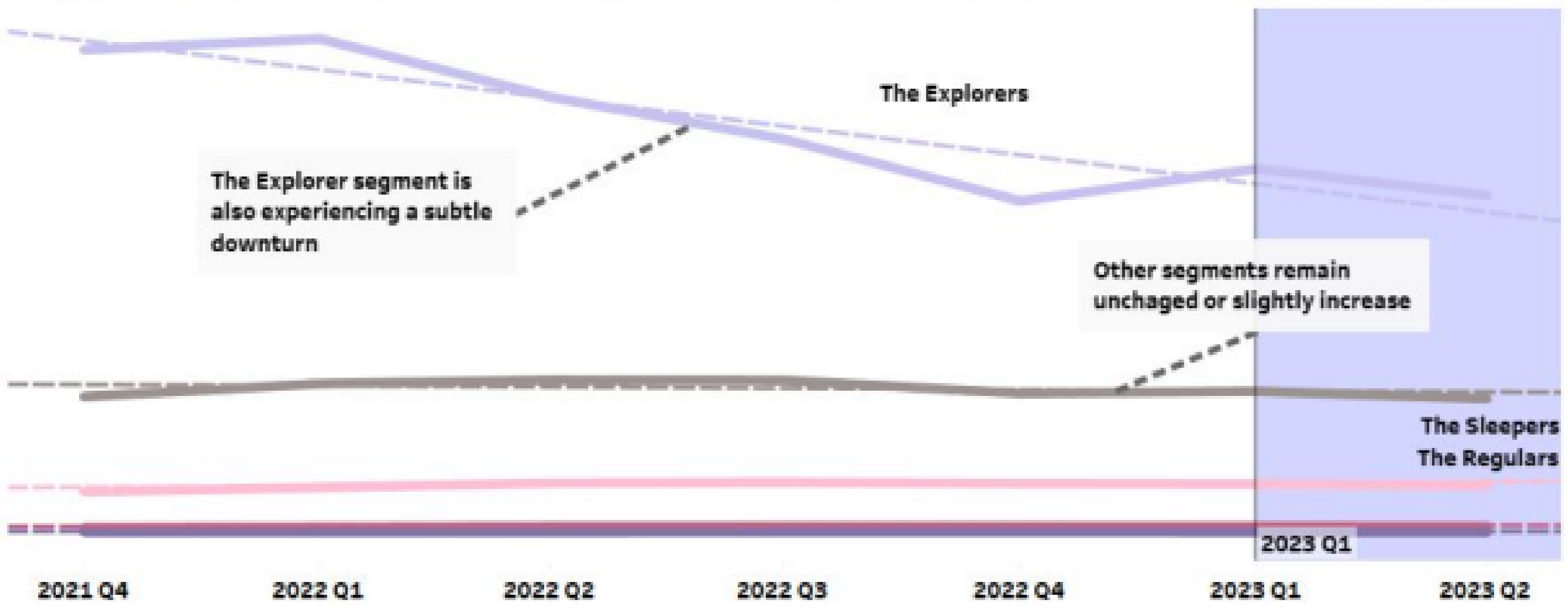
Full Slides Here

The decline in customer acquisition is largely influenced by a gradual reduction in The Explorers segment. Although reducing in size, each customers from this group seem to be more valuable over time.

Total number of active customers is facing a gradual reduction...

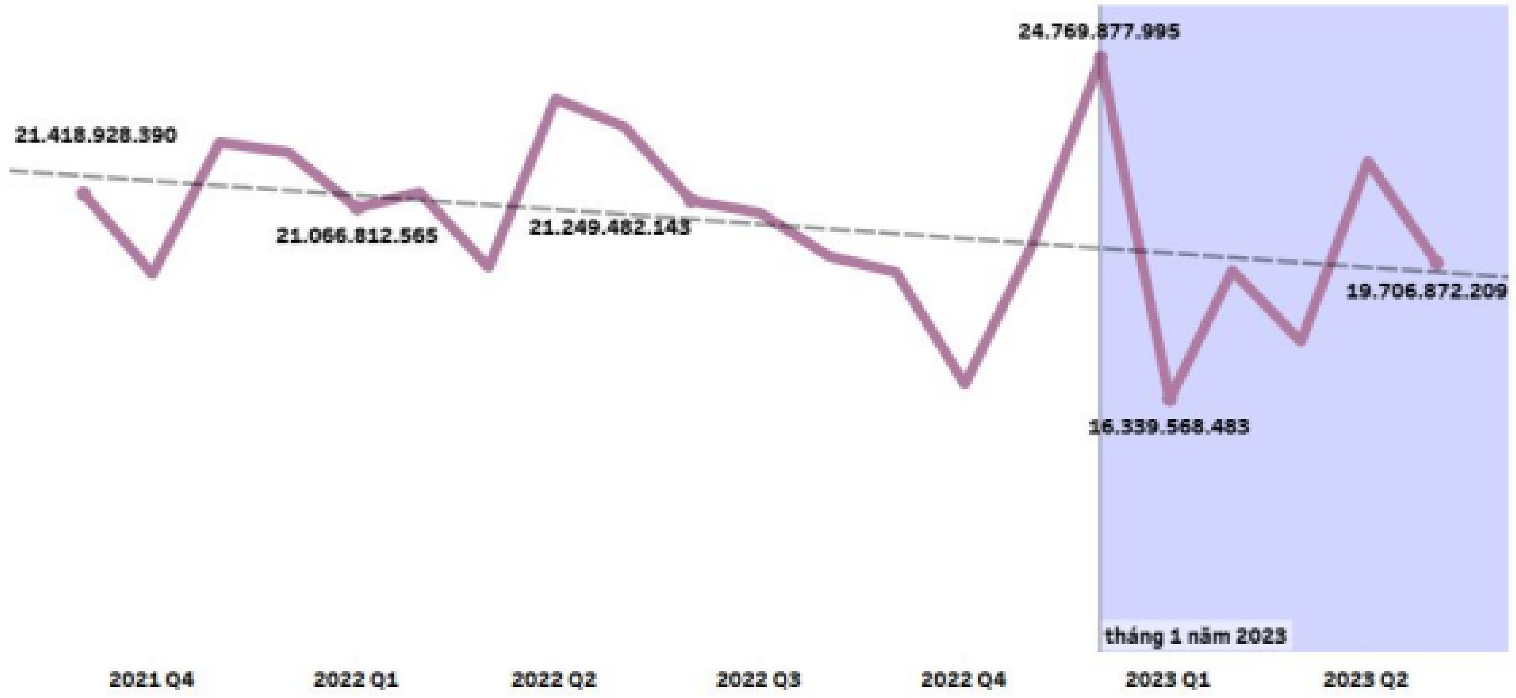


...which is drived by the steady decrease in the largest segment: The Explorer

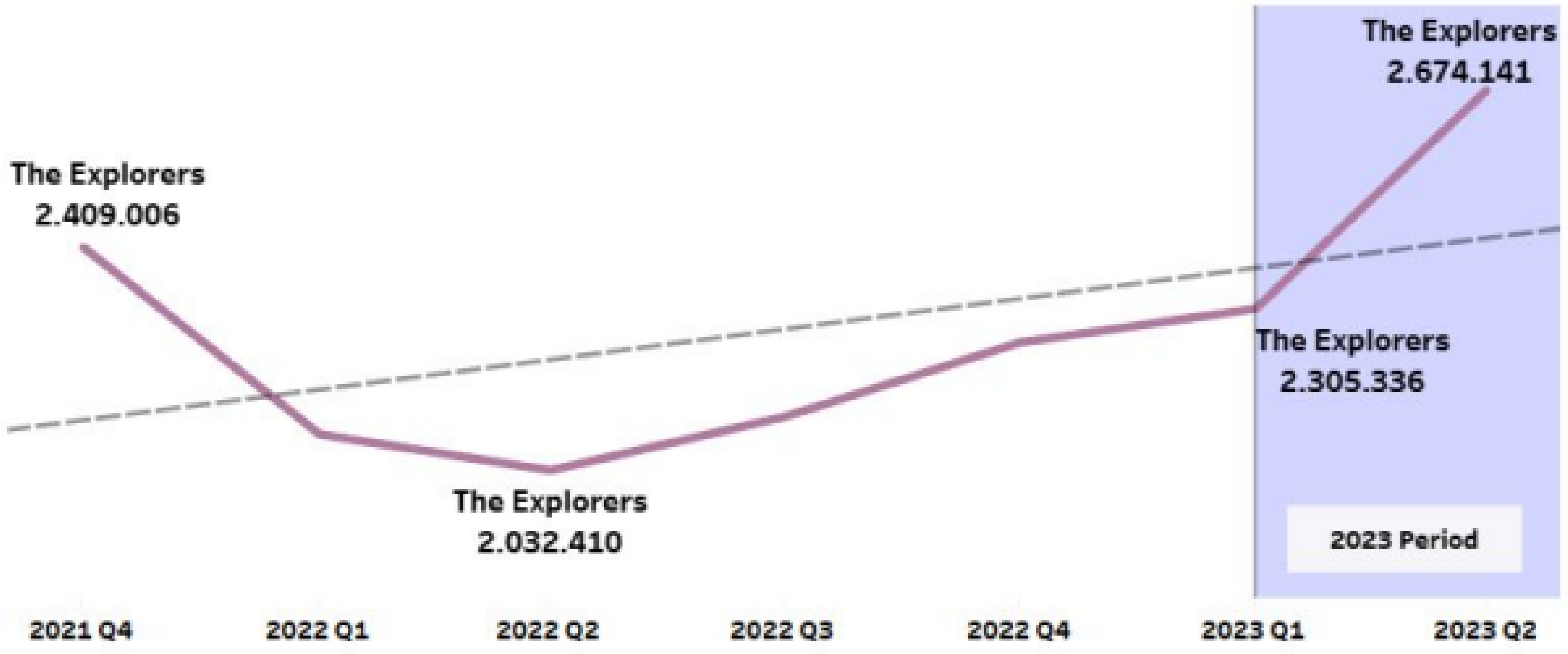


*This slide is an example

... Had led to a declining pattern in sales



However, CLV of Explorers are forcasted to continue growing, indicating a more loyal segment.



project

PORTFOLIO

01 Pizza Hut Analysis

Analyzing and Story-telling
"How to grow market?"

02 Brazilian E-commerce Analysis

Extracting and visualizing
Advanced Tableau - Python

03 Other projects

5+ Projects from Basic
analytics to Deep learning

BRAZILIAN E-COMMERCE

Link to final product: Tableau Dashboard | Kaggle Notebook

PROJECT GOAL

Extracting management insights with Python and presenting outcomes through a Tableau **dashboard** covering **Customer Segments, Sales, and Operations.**

SKILLS

- **Data visualization**
- **Interactive Dashboard making**
- **Sale, Sentiment, Operation analytics**
- **Business Acumen**

KEY TASKS

- Generated an interactive **dashboard for sales** in Tableau.
- Calculated 5 **operational metrics**, such as order processing speed and late shipment days.
- Revealed **bought-together items** (used for a recommendation system).
- Performed **sentiment analysis** to identify the most favored products.
- Calculated 4 customer metrics, such as **comment ratio** and value per order.
- Classified the customer base into 10 segments using the **RFM framework.**

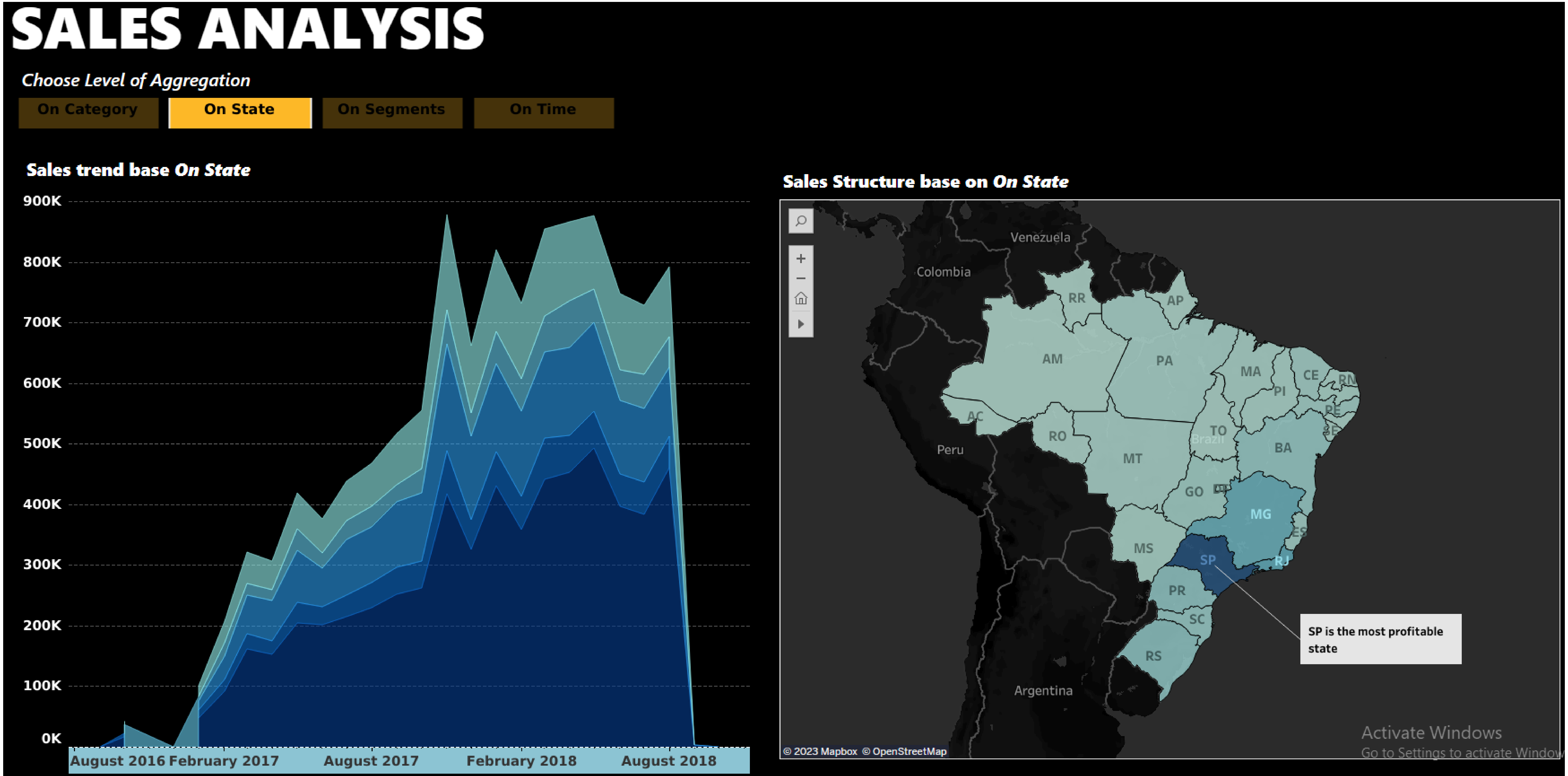
TOOLS

- **Python:** Pandas, Matplotlib,...
- **Tableau:** Interactive Dashboard

E-COMMERCE

GENERATED AN INTERACTIVE DASHBOARD FOR SALES IN TABLEAU

DashBoard Here



E-COMMERCE

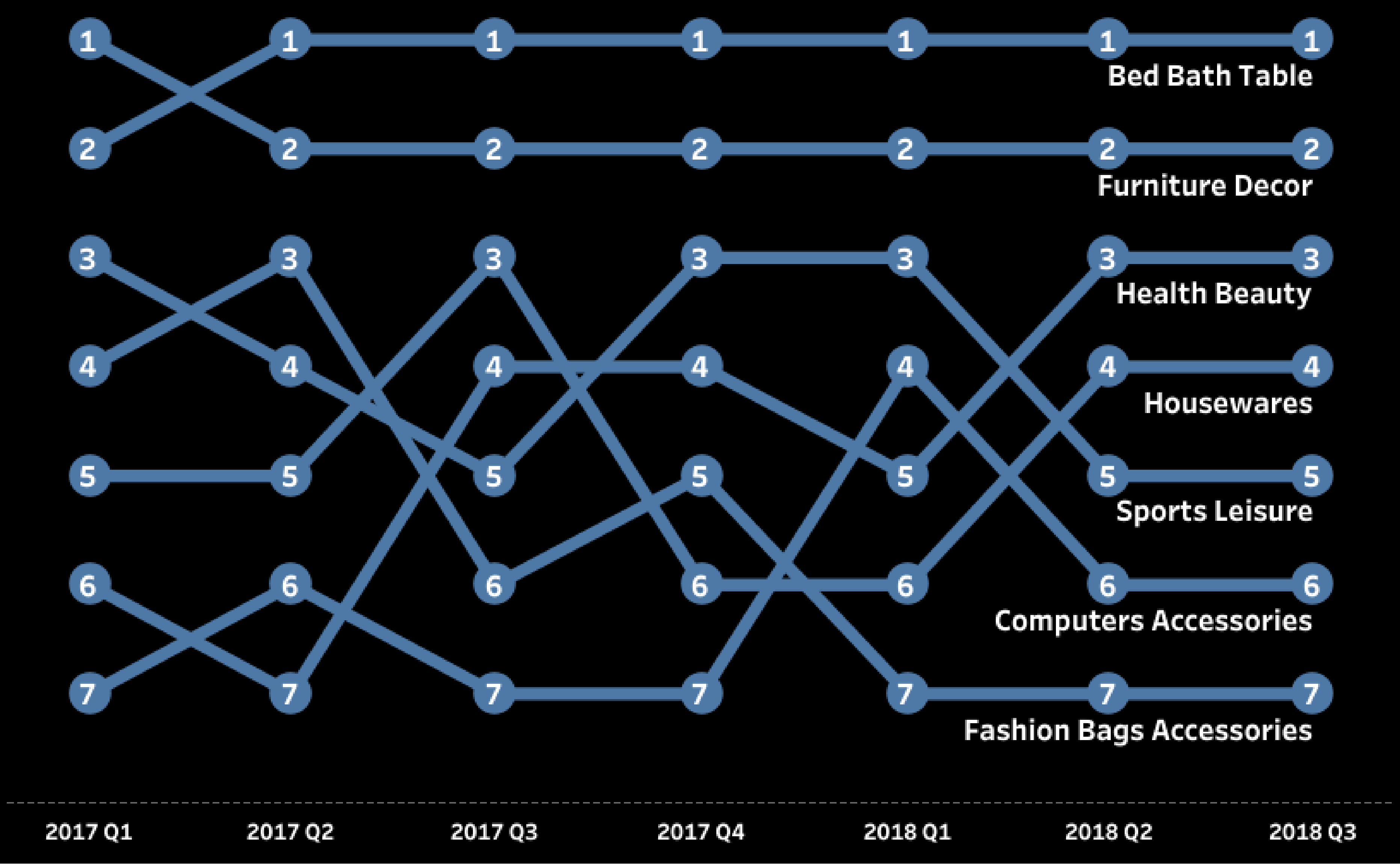
REVEALED MOST FAVORED PRODUCTS || BOUGHT-TOGETHER ITEMS

DashBoard Here

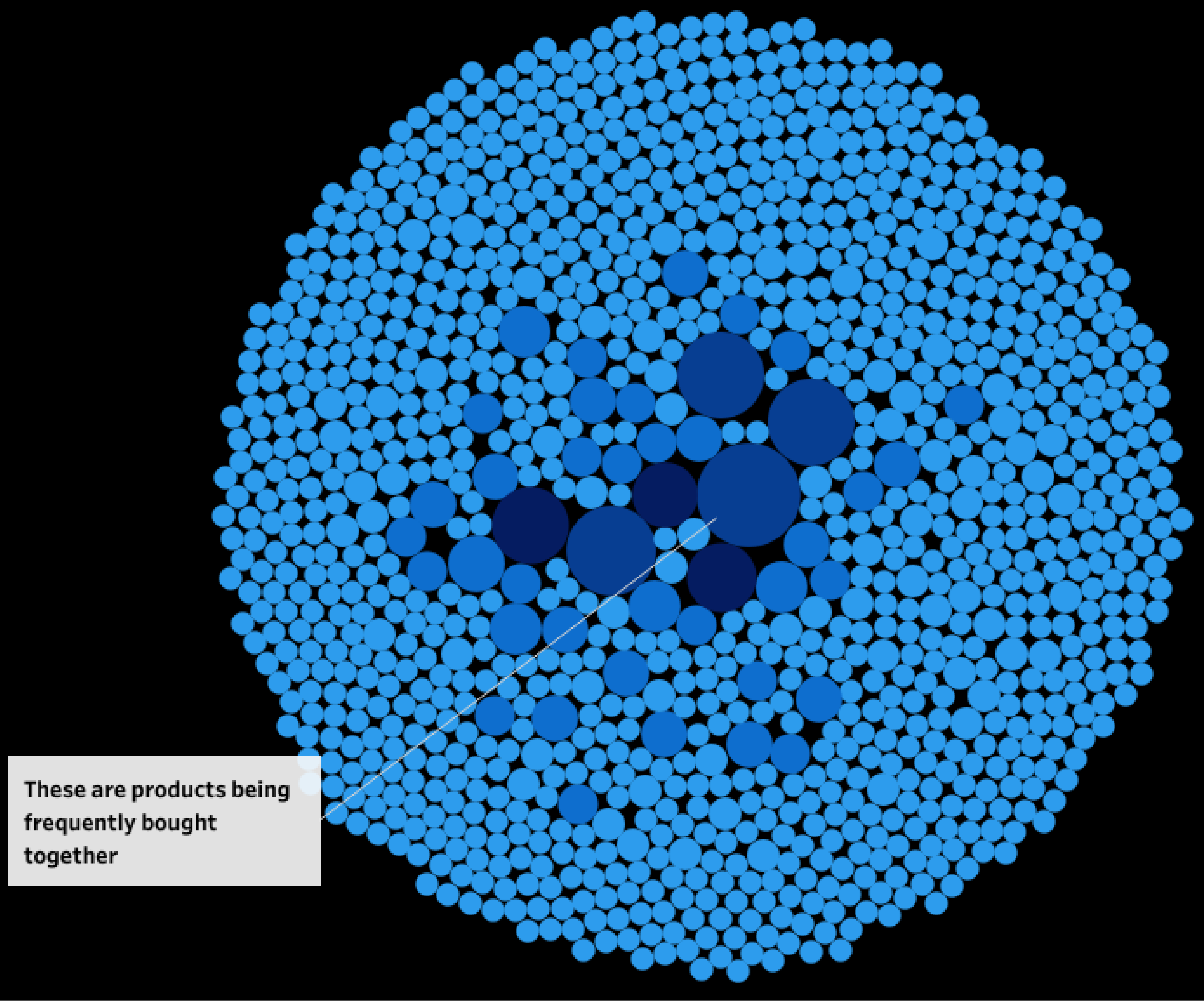
PRODUCT ANALYSIS

Product rank change over time

Rank is assigned base on number of purchase | Click on the category name to highlight path

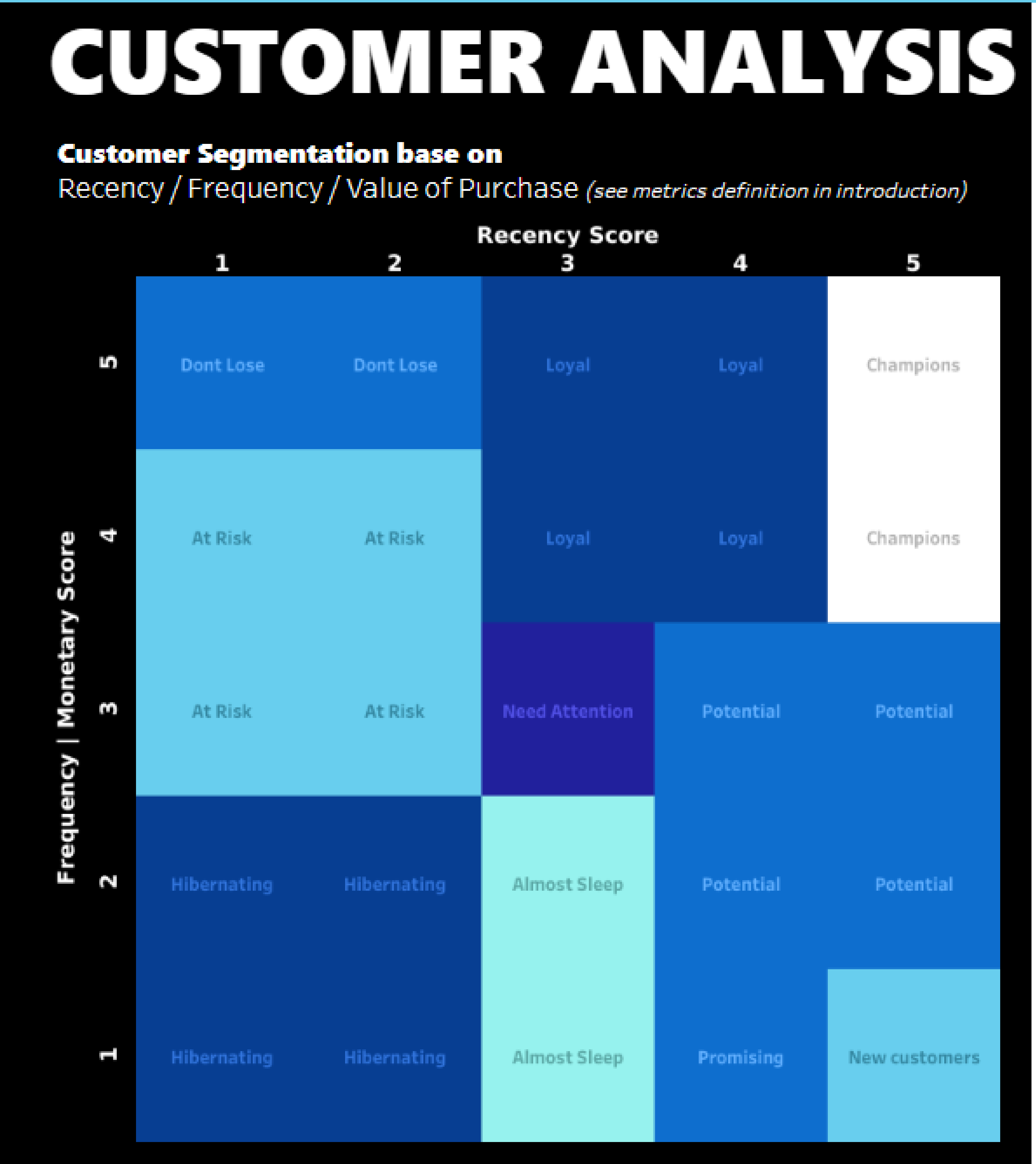


Map of Products that are bought together



E-COMMERCE

CLASSIFIED THE CUSTOMER BASE USING THE RFM FRAMEWORK



1 | RFM ANALYSIS

DashBoard Here
Notebook Here

1.1. RECENCY

```
# getting the latest day of purchase for each user
recency_table = customer.merge(order, on ="customer_id")
recency_table = recency_table.groupby('customer_unique_id').agg({'order_purchase_timestamp': 'ma
x'}).reset_index()
# convert to datetime dtype
recency_table['order_purchase_timestamp'] = pd.to_datetime(recency_table['order_purchase_timestam
p'])
# calculate the current day
current_day = pd.to_datetime(max(order['order_purchase_timestamp']))
# calculate the days between
recency_table['days_between'] = (current_day - recency_table['order_purchase_timestamp']).apply(lamb
da x: int(str(x).split()[0]))
# def function that scores the data
def scoring(x):
    # calculate quantiles
    quan_val_list = []
    for quan in [0.2,0.4,0.6,0.8]:
        quan_val_list.append(np.quantile(x,quan))
    # label the data base on the quantile value
    def labeling(x):
        if x < quan_val_list[0]: return 1
        elif x < quan_val_list[1]: return 2
        elif x < quan_val_list[2]: return 3
        elif x < quan_val_list[3]: return 4
        else: return 5

    return x.apply(labeling)

recency_table['recency_score'] = scoring(recency_table['days_between'])
recency_table
```

project PORTFOLIO

01 Pizza Hut Analysis

Analyzing and Story-telling
"How to grow market?"

02 Brazilian E-commerce Analysis

Extracting and visualizing
Advanced Tableau - Python

03 Other projects

5+ Projects from Basic analytics to Deep learning

PROJECTS	OUTCOME	DESCRIPTION
Statistical Analysis Dognition: 80% Of Stats You Need	**Kaggle notebook:** 80% of Statsitics you need_Dognition data | Kaggle	**Apply 80% of statistical techniques to the Dognition dataset, creating a guidance notebook for community use.** • **Skills/Tools:** Statistical methods | Python programming • **Key Tasks:** Executed Sampling and Examining Normality | Descriptive Analysis | Basic & Advanced Hypothesis Testing | ootstrapping and Permutation
A/B Testing: Ads displayment policy of Cookie Cat's Mobile game	**Kaggle notebook**: A/B testing | Kaggle	**Employ statistical methods to investigate the influence of Cookie Cats' new ad display policy on player retention.** • **Skills/Tools:** A/B Testing | Statistical methods • **Key Tasks:** Performed permutation testing - a non-parametric statistical approach
Analyzing HR data using logistic regression, decision tree	**Kaggle notebook:** HR Analytics | Kaggle	**Reveal factors influencing a company's attrition rate (employee turnover) using logistic regression and decision tree analysis.** • **Skills/Tools:** Behaviour anlysis | HR | Statistical methods | Python • **Key Tasks:** Applied logistic regression and decision tree analysis, comparing variable coefficients and feature importance to make final recommendations.

PROJECTS	OUTCOME	DESCRIPTION
Financial Statement Dashboard with Tableau	**Tableau:** Financial Statement Dashboard | Tableau Public	**Visualized the key financial information for SP500 companies.** • **Skills/Tools:** Financial analysis | Tableau | Interactive Dashboard • **Key Tasks:** Visualized balance sheet | income funnel | financial ratios
Disasters classification with RNN - LSTM - BERT	**Kaggle notebook**: LSTM-BERT-TweetData | Kaggle	**Train and evaluate some NLP models to help classify if a tweet is about a real disaster or not.** • **Skills/Tools:** NLP | Deep Learning | Tensorflow | BERT • **Key Tasks:** Build, train, evaluate Simple RNN | LSTM | BERT
Churn rate prediction with XGB - Neural network	**Kaggle notebook:** Churn predict XGB NeuNet | Kaggle	**Predict customer churn in a telecom company.** • **Skills/Tools:** XGB • **Key Tasks:** Training and evaluating XGB model.

THANKS FOR WATCHING

Portfolio with All projects:
Here

Linkedin:
Here