

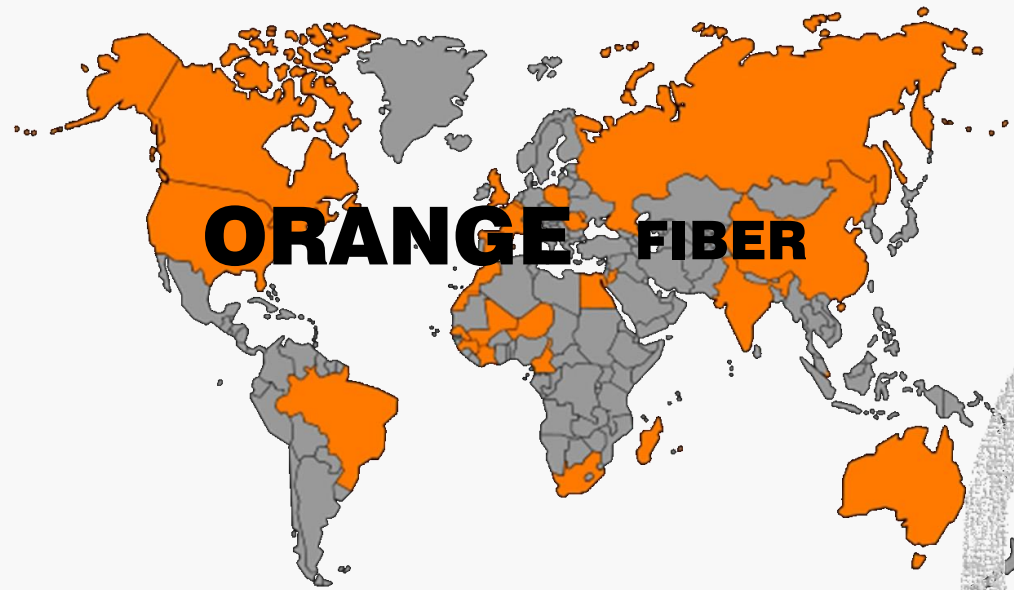


orangeTM



ORANGE FIBER

orangeTM



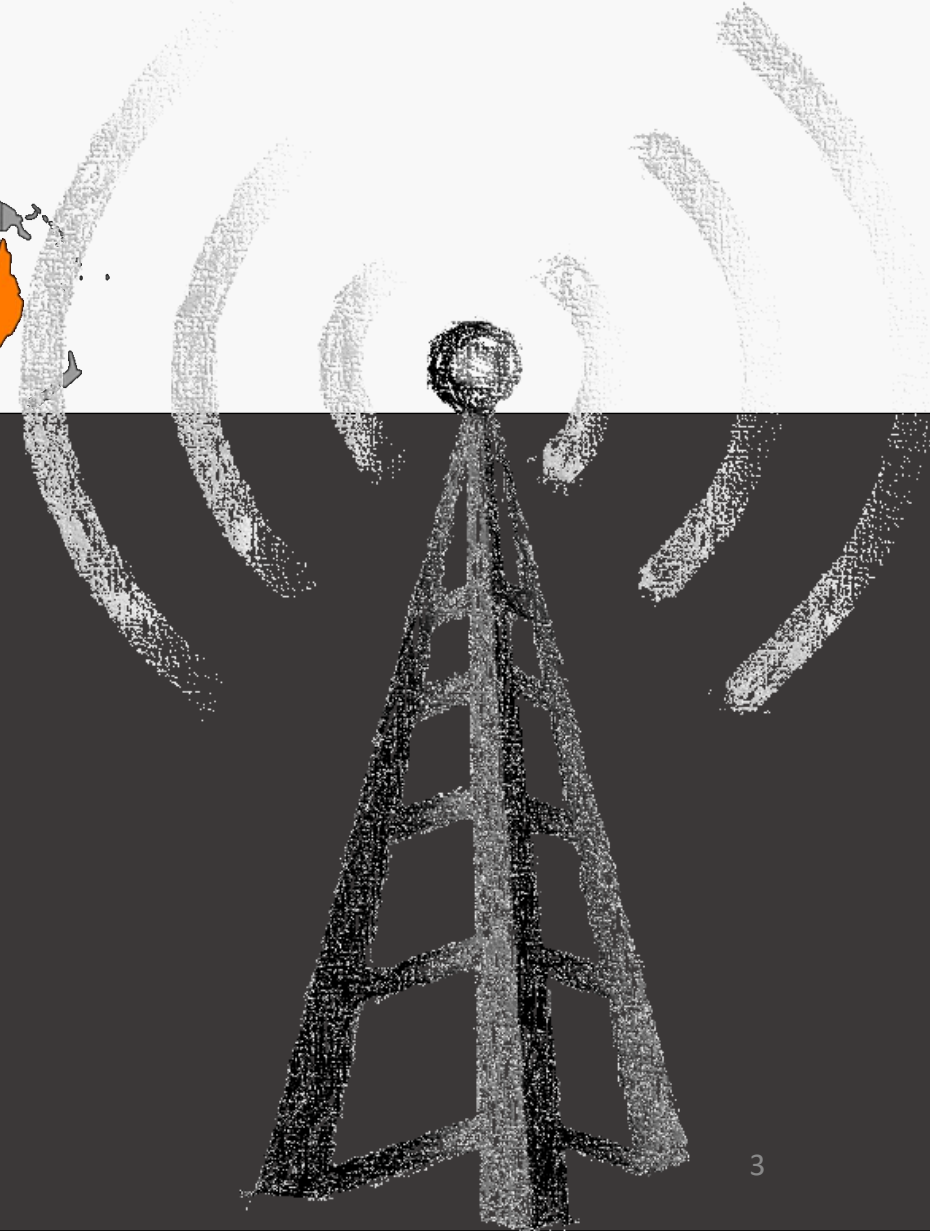
CODING ACADEMY BY ORANGE

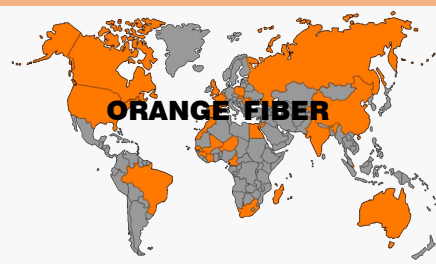
TASK 1

RAZAN SALMAN

NEDAL ALTITI

LAITH RASHEED





O U T L I N E S

PROBLEM DEFINITION

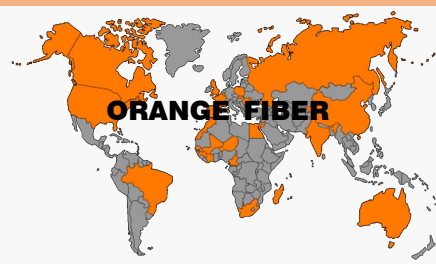
PROBLEM STRUCTURE

FEATURES

HYPOTHESIS

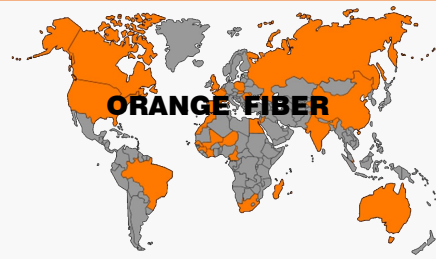
RECOMMENDATION





PROBLEM DEFINITION



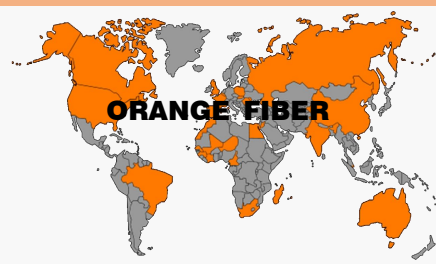


PROBLEM DEFINITION

INITIAL SITUATION

Dataset that includes details about fiber customers, demographics, subscription technical information, and contract type





PROBLEM DEFINITION

PROJECT SCOPE/ USE CASE

1. OUR OBJECTIVE

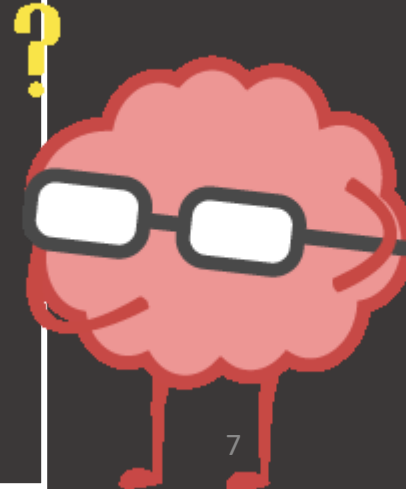
2. BUISSNIS OBJECTIVE

OUR OBJECTIVE

identify customers who are at risk of churning and check their loyalty and satisfaction

BUISSNES OBJECTIVE

REDUCE THE CHURN RATE
AND IMPROVING CUSTOMER
RETENTION



What is the churn rate?

“Churn rate is a metric used to measure the rate at which customers or subscribers discontinue using a company's products or services over a given period of time”



What is the churn rate?

“Churn rate is a metric used to measure the rate at which customers or subscribers discontinue using a company's products or services over a given period of time”

REASONS

- 1- CUSTOMER DIES OR GOES OUT OF BUSINESS.
- 2- JUST DO NOT ACHIEVE THEIR DESIRED OUTCOME.



What is the churn rate?

“Churn rate is a metric used to measure the rate at which customers or subscribers discontinue using a company's products or services over a given period of time”

CHURN RATE CALCULATION

CHURN RATE CAPTURES THE NUMBER OF PEOPLE WE RETAIN AT THE END OF A TIME PERIOD.

$$\text{CHURN RATE} = \frac{\text{CUSTOMERS AT START OF USAGE INTERVAL} - \text{CUSTOMERS AT END OF USAGE INTERVAL}}{\text{CUSTOMERS AT START OF USAGE INTERVAL}}$$

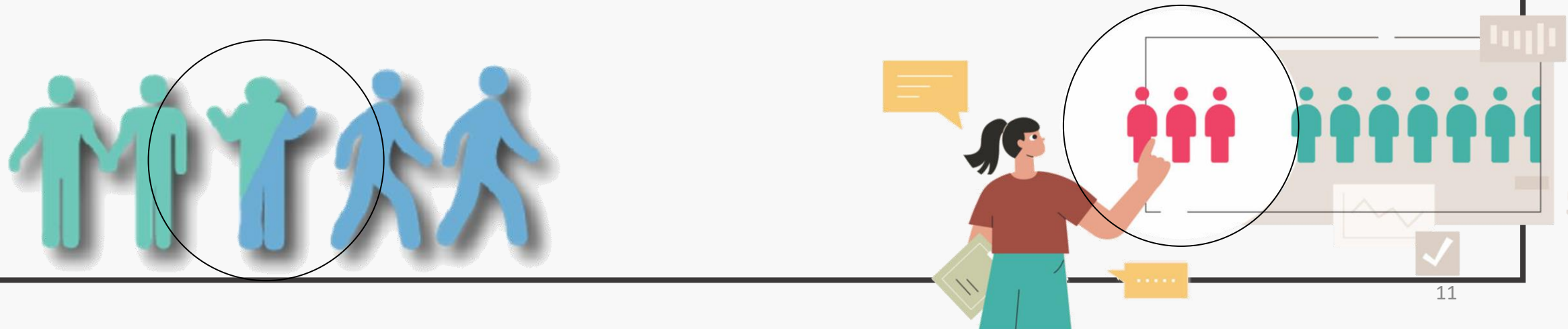


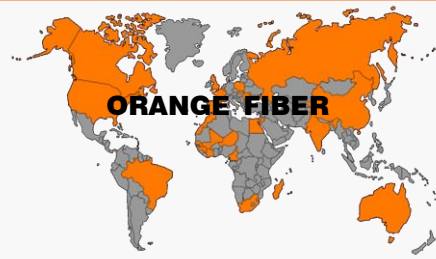
What is the churn rate?

“Churn rate is a metric used to measure the rate at which customers or subscribers discontinue using a company's products or services over a given period of time”

SOLUTIONS

identify customers who are at risk of churning and take steps to retain them.



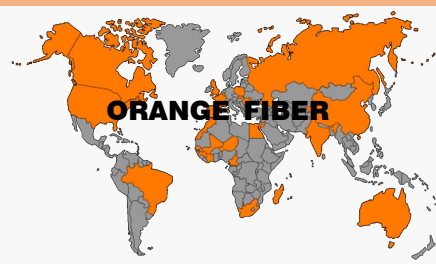


PROBLEM DEFINITION

SUCCESS CRITERIA

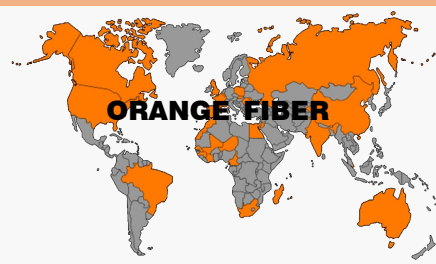
in general, an **annual** churn rate of 5% is seen as a reasonable benchmark





PROBLEM STRUCTURE





PROBLEM STRUCTURE

“WHAT COULD BE THE KEY ELEMENT OF THE PROBLEM?”

FRAME THE PROBLEM

GET THE DATA

EDA

**MISINTERPR
ETATION
SOME
FEATURES**

**MISSING
VALUES**

**OUTLIERS
AND
VARIATION**

**IMBALANCE
DATA**

**CATEGOR-
ICAL DATA**

PROBLEM STRUCTURE

“WHAT COULD BE THE KEY ELEMENT OF THE PROBLEM?”

FRAME THE PROBLEM

GET THE DATA

EDA

**MISINTERPR
ETATION
SOME
FEATURES**

- 1. SEARCH**
- 2. ASK EXPERTS**
- 3. CASE STUDIES**

**MISSING
VALUES**

- 1. Structured**
- 2. MAR**
- 3. MNAR**

**OUTLIERS
AND
VARIATION**

- 1. Transforma
tion**
- 2. Modeling**
- 3-
Normalization**

**IMBALANCE
DATA**

- 1. Non
represent
ative**
- 2. skewness**

**CATEGOR-
ICAL DATA**

- 1. Encoding
to convert
them to
numerical
values**

PROBLEM STRUCTURE

“WHAT COULD BE THE KEY ELEMENT OF THE PROBLEM?”

MISSING VALUES

	columns	number of missing values
0	OF_PREV_SPEED	74714
1	GB_TOTAL_CONSUMPTION_Month3	4242
2	Disconnection_TOTAL_MAX_day	625
3	Disconnection_TOTAL_MIN_day	625
4	Disconnection_TOTAL_SUM_Month	625
5	Disconnection_TOTAL_MEAN_Month	625
6	GB_TOTAL_CONSUMPTION_Month2	367
7	GOVERNORATE	323
8	GB_TOTAL_CONSUMPTION_Month1	155
9	CUSTOMER_GENDER	54
10	LAST_LINK_QUALITY	28
11	LAST_LINK_STATUS	28
12	LAST_POWER_VALIDATION	28
13	LAST_LINK_PRIORITY	28

We have 82,467 missing values in total

PROBLEM STRUCTURE

“WHAT COULD BE THE KEY ELEMENT OF THE PROBLEM?”

MISSING VALUES

Methods to handle the missing values

Numerical Variable

- 1- delete the missing values x**
- 2- using the mean x**
- 3- using the median, better for outlier and skewness**
- 4- iterative imputation**

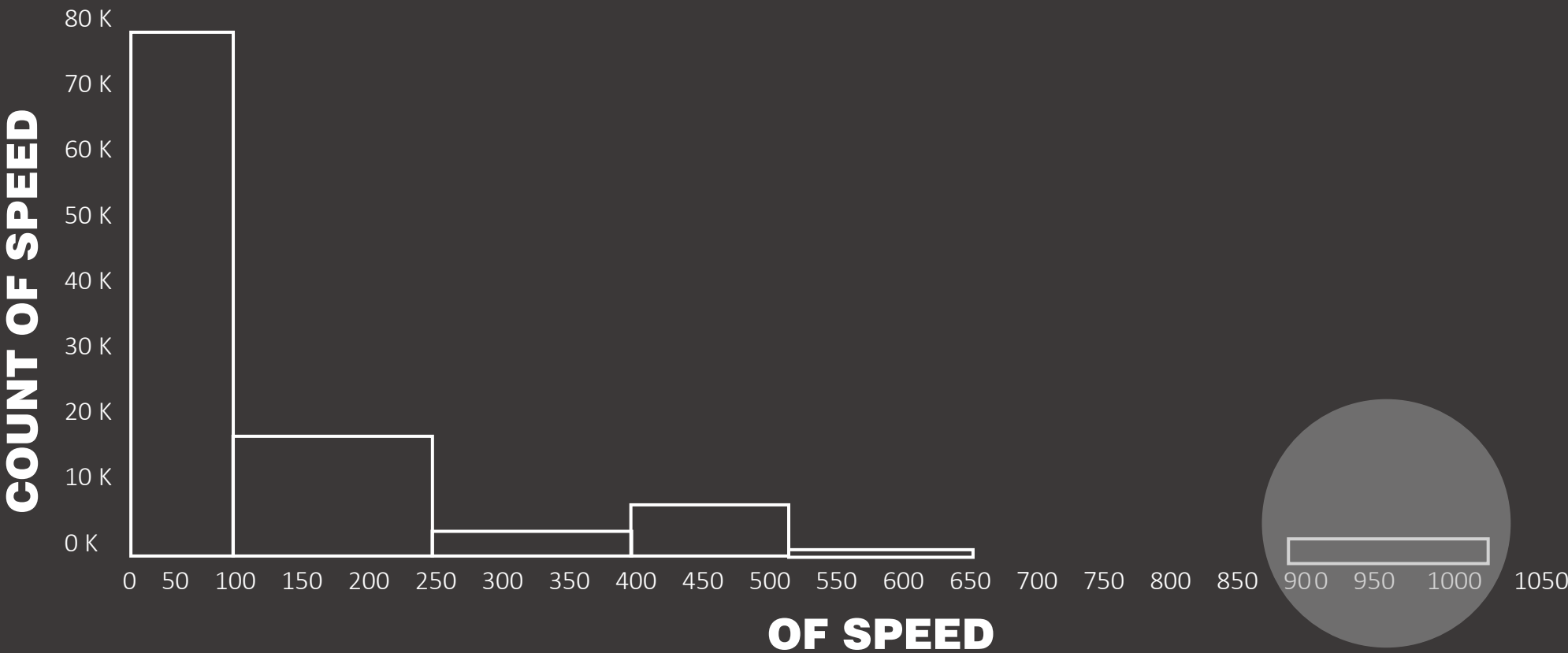
Categorical Variable

- 1- delete the missing values**
- 2- impute the missing values with most frequent variable**
- 3- iterative imputation**

PROBLEM STRUCTURE

“WHAT COULD BE THE KEY ELEMENT OF THE PROBLEM?”

Outliers

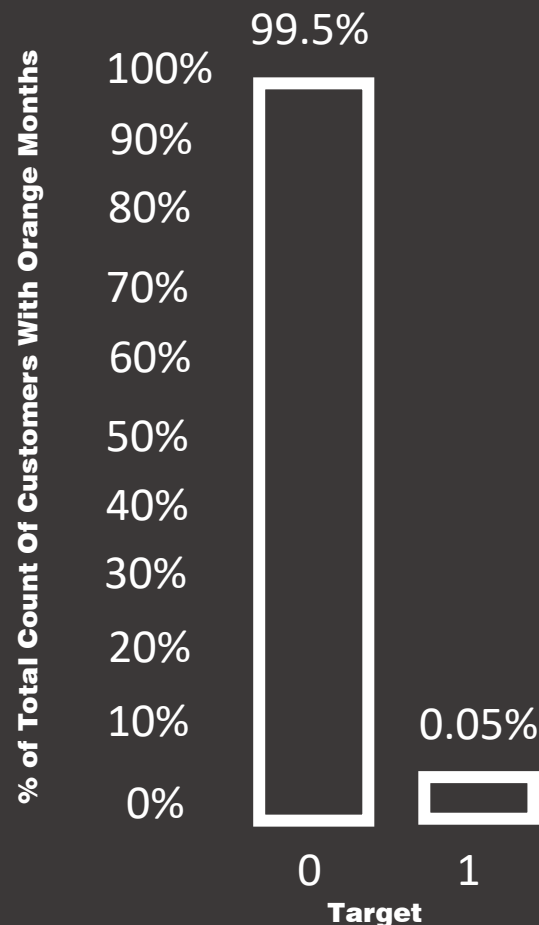


PROBLEM STRUCTURE

“WHAT COULD BE THE KEY ELEMENT OF THE PROBLEM?”

IMBALANCE DATA

CHURNED vs UNCHARNED CUSTOMERS



Target 0 : people who Stay with Orange

Target 1 : People who left Orange

Methods to handle the imbalance data

1- Combine oversampling and undersampling and keep in mind to take the sample that is representative to the population

2- Oversampling synthetic Data

3-Use algorithms with weighted average

PROBLEM STRUCTURE

“WHAT COULD BE THE KEY ELEMENT OF THE PROBLEM?”

Categorical data

Methods to handle categorical data

1- OneHotEncoding, e.g “Gender, Migration Flag”.

2- LabelEncoding, e.g “age group”.

OneHotEncoding

TARGET	F	M
0	0	1
1	1	0

LabelEncoding

AGE GROUP	CATEGORICAL	TARGET
CHILD	1	1
TENEEAGER	2	1
ADULT	3	0
SENIOR	4	0

PROBLEM STRUCTURE

“WHAT COULD BE THE KEY ELEMENT OF THE PROBLEM?”

FRAME THE PROBLEM

GET THE DATA

EDA

**MISINTERPR
ETATION
SOME
FEATURES**

- 1. SEARCH**
- 2. ASK EXPERTS**
- 3. CASE STUDIES**

**MISSING
VALUES**

- 1. Structured**
- 2. MAR**
- 3. MNAR**

**OUTLIERS
AND
VARIATION**

- 1. Transforma
tion**
- 2. Modeling**
- 3-
Normalization**

**IMBALANCE
DATA**

- 1. Non
represent
ative**
- 2. skewness**

**CATEGOR-
ICAL DATA**

- 1. Encoding
to convert
them to
numerical
values**

PROBLEM STRUCTURE

“WHAT COULD BE THE KEY ELEMENT OF THE PROBLEM?”

FRAME THE PROBLEM

GET THE DATA

EDA

**MISINTERPR
ETATION
SOME
FEATURES**

**MISSING
VALUES**

**OUTLIERS
AND
VARIATION**

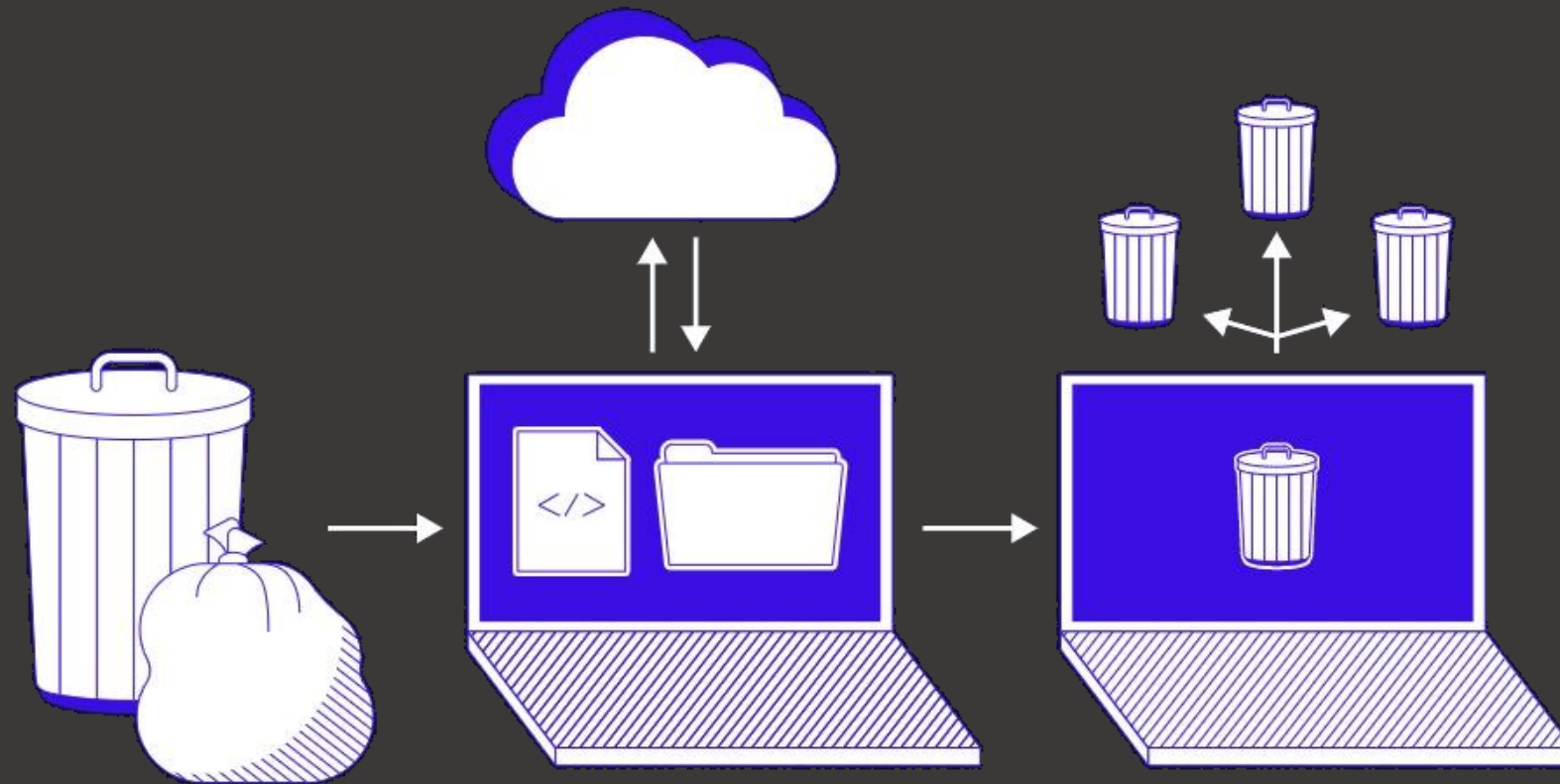
**IMBALANCE
DATA**

**CATEGOR-
ICAL DATA**

DATA GAB

DATA GAB

GARBAGE IN, GARBAGE OUT



Bad Data

Great Model

Bad Prediction

FRAME THE PROBLEM

GET THE DATA

EDA

**MISINTERPR
ETATION
SOME
FEATURES**

**MISSING
VALUES**

**OUTLIERS
AND
VARIATION**

**IMBALANCE
DATA**

**CATEGOR-
ICAL DATA**

DATA GAB

**CODING
ACADEMY
BY ORANGE
TASK 1**

FRAME THE PROBLEM

GET THE DATA

EDA

DATA GAB

VISUALIZATION

**EXPLORE DIFFERENT
MODELS**

**FINE TUNNING THE
MODEL**

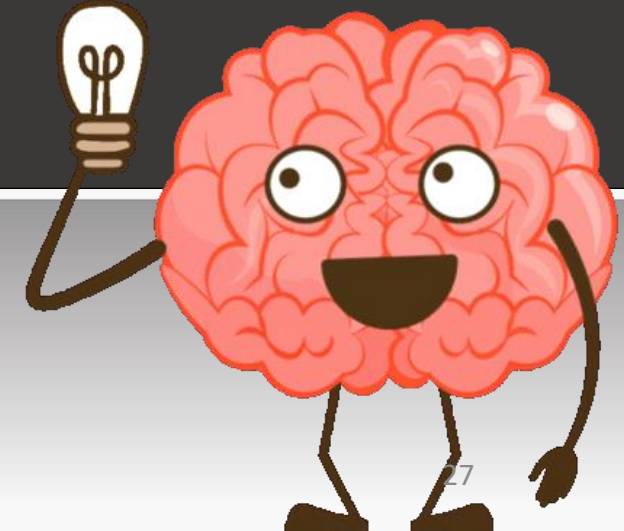
EVALUATION METRICS

**COMBINING PRECISION AND RECALL, F1 SCORE, AREA UNDER CURVE
ROC**

PRESENT SOLUTION

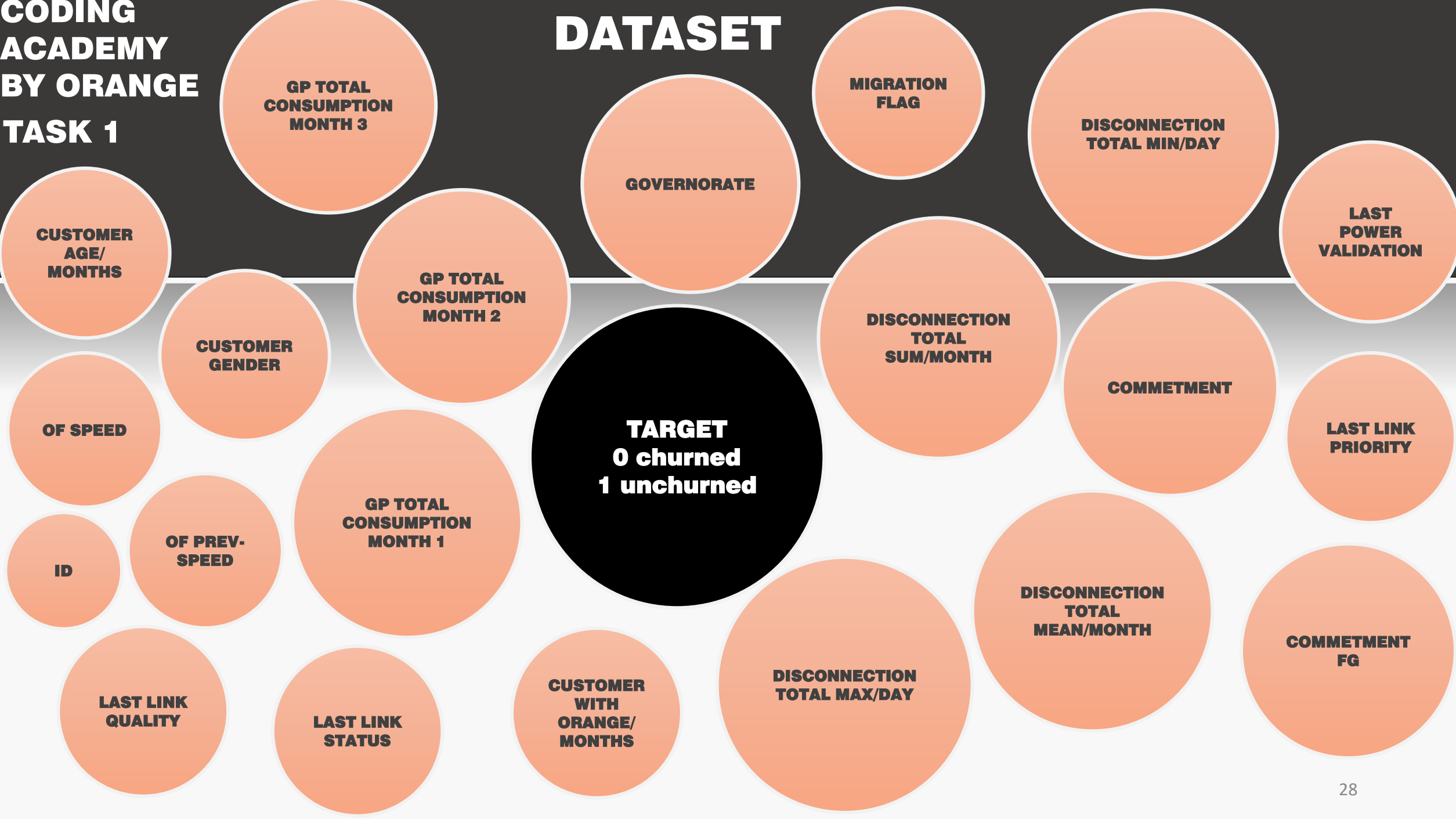
DEPLOY THE MODEL



[illegible]

**CODING
ACADEMY
BY ORANGE
TASK 1**

DATASET

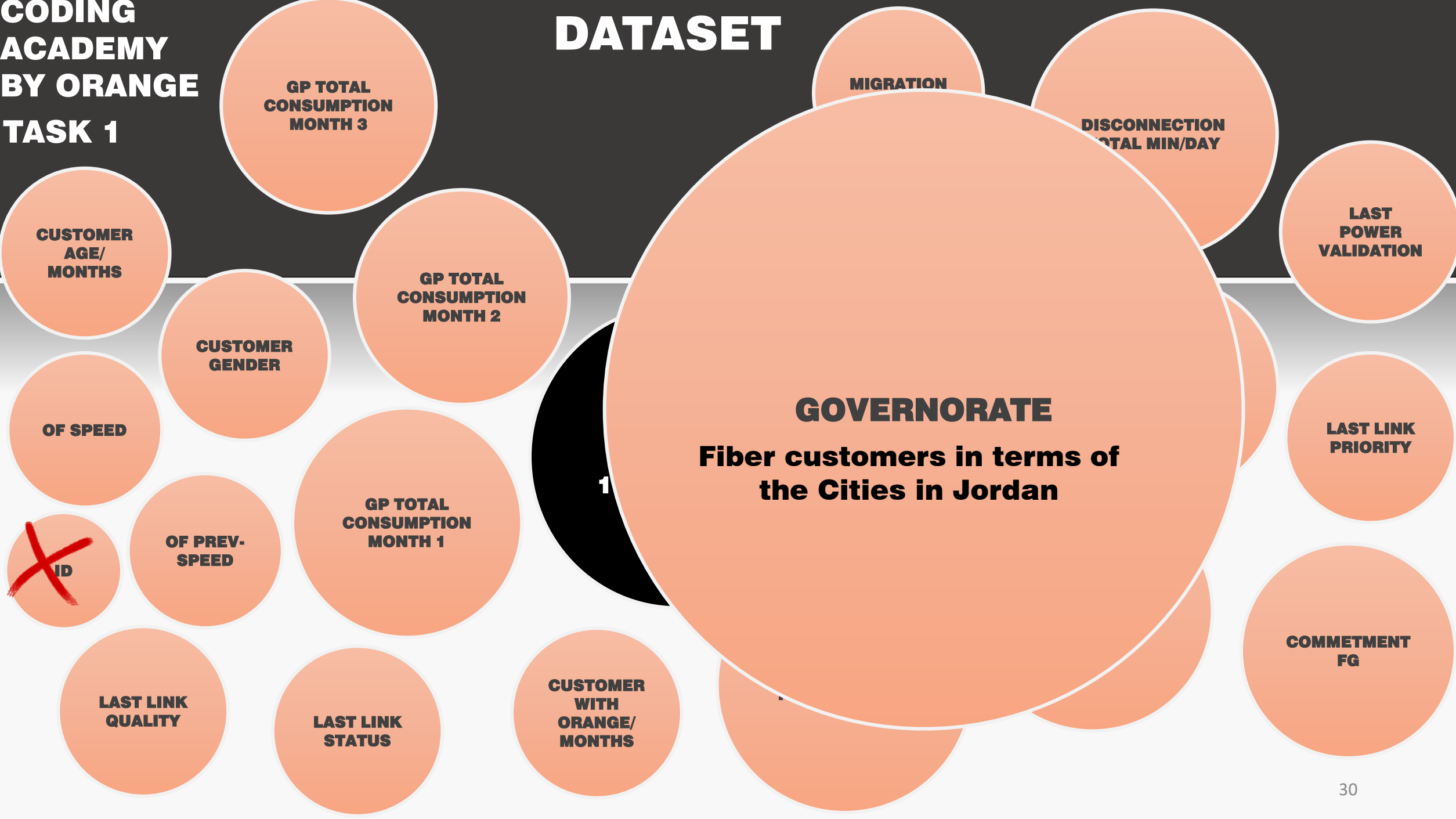


DATASET

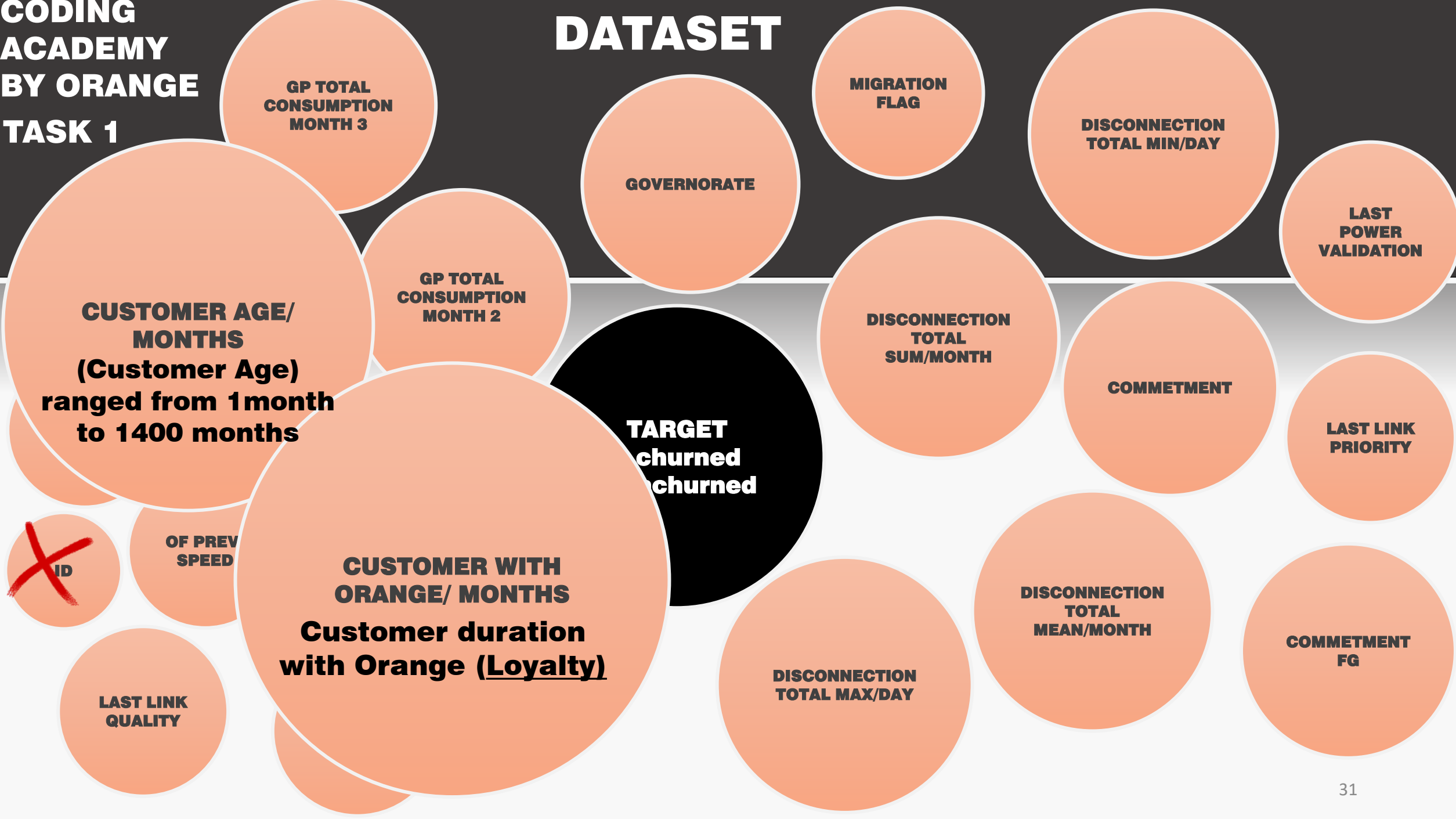


**CODING
ACADEMY
BY ORANGE
TASK 1**

DATASET



DATASET



DATASET

**CUSTOMER AGE/
MONTHS
(Customer Age)
ranged from 1month
to 1400 months**

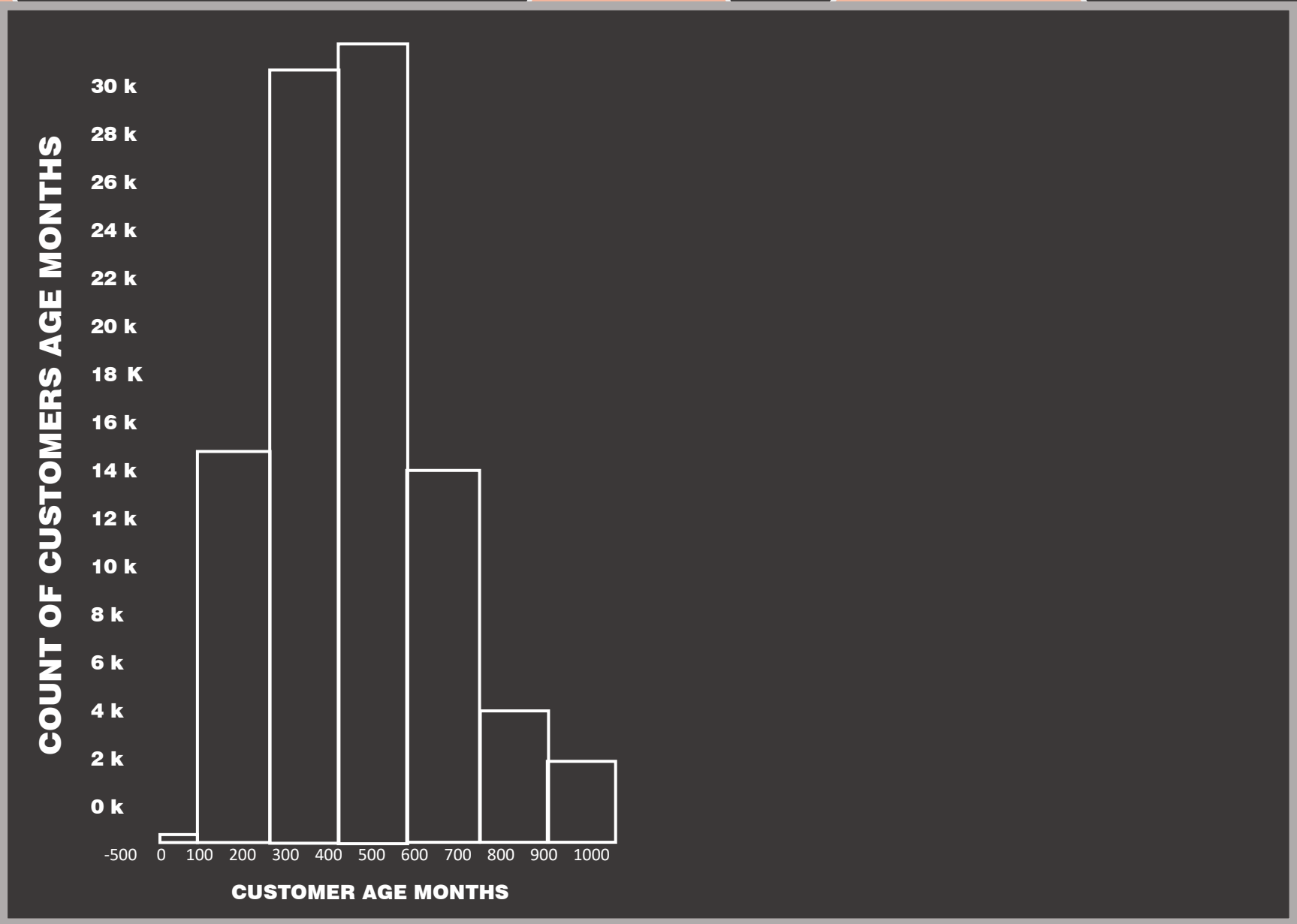
~~ID~~

**OF PREV-
SPEED**

**GP TO
CONSUM
MON'**

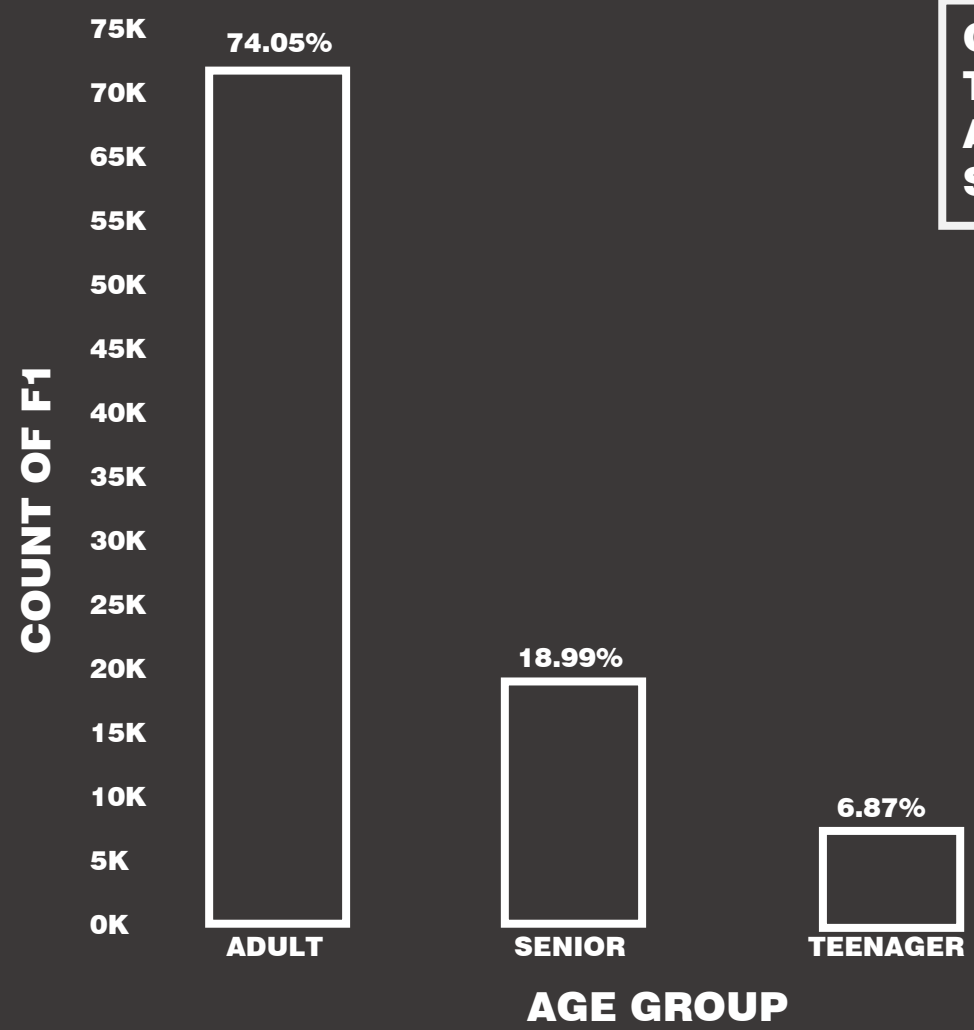
**LAST LINK
QUALITY**

**LAST LINK
STATUS**

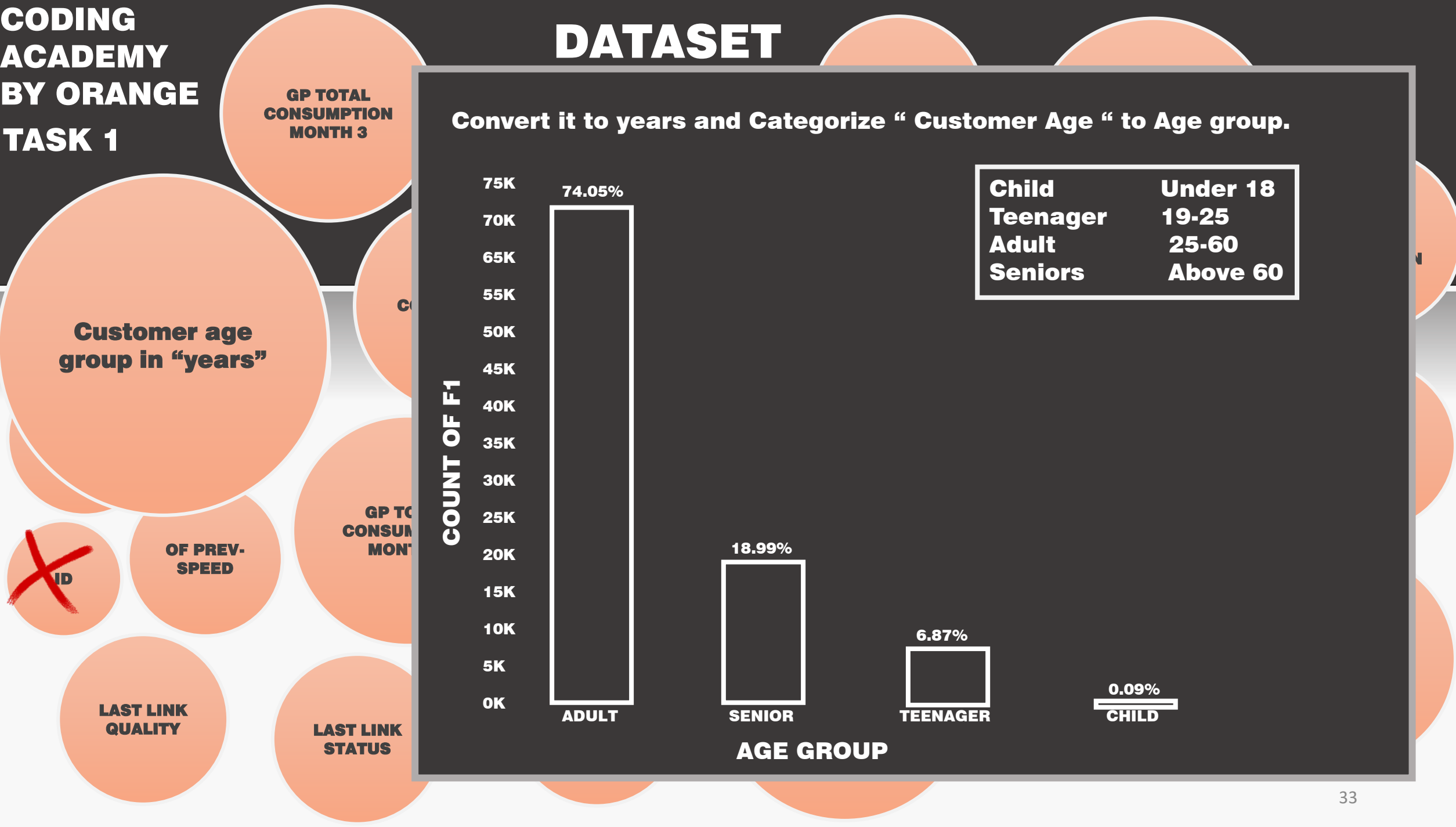


DATASET

Convert it to years and Categorize “ Customer Age “ to Age group.

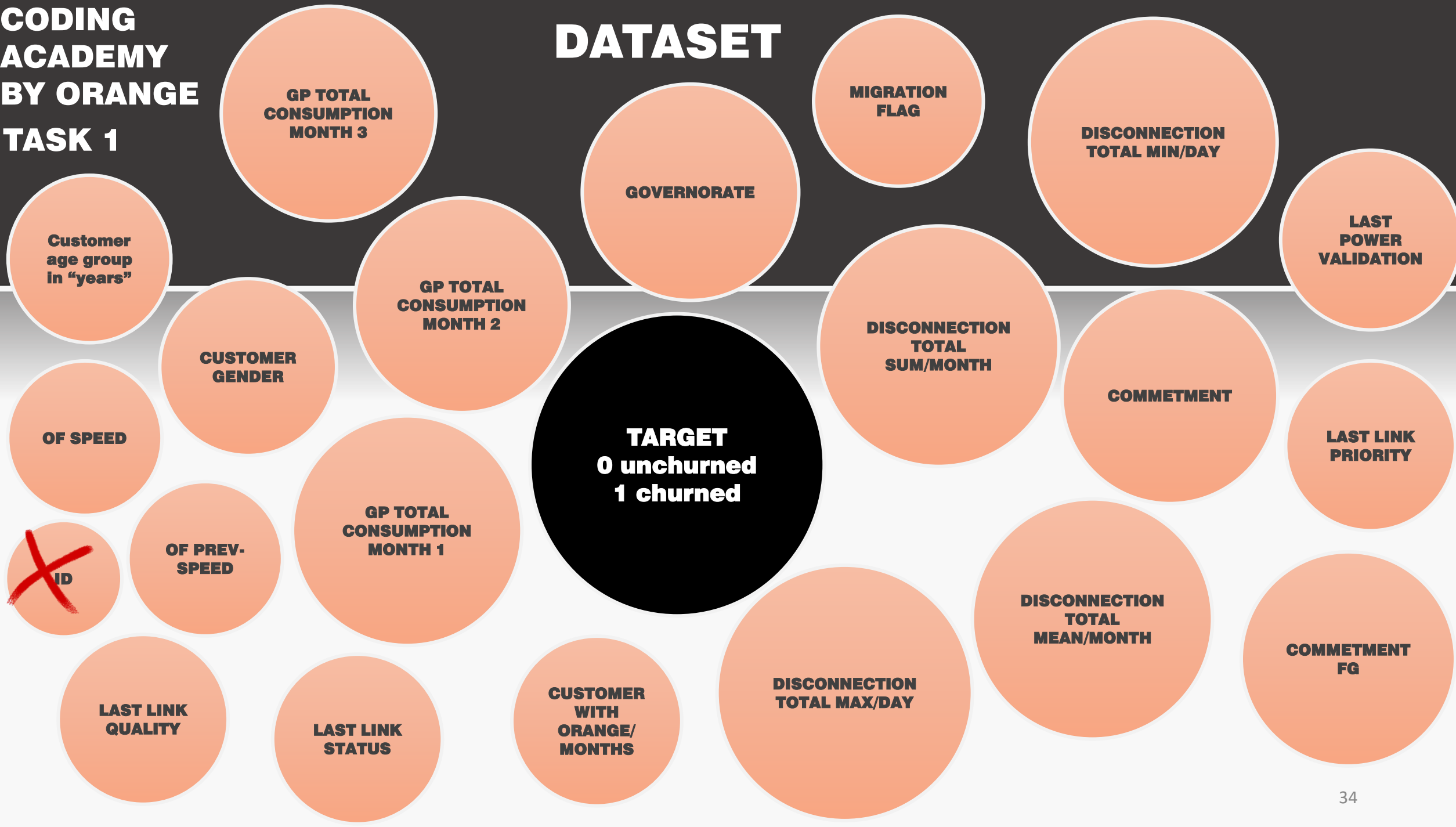


Child	Under 18
Teenager	19-25
Adult	25-60
Seniors	Above 60



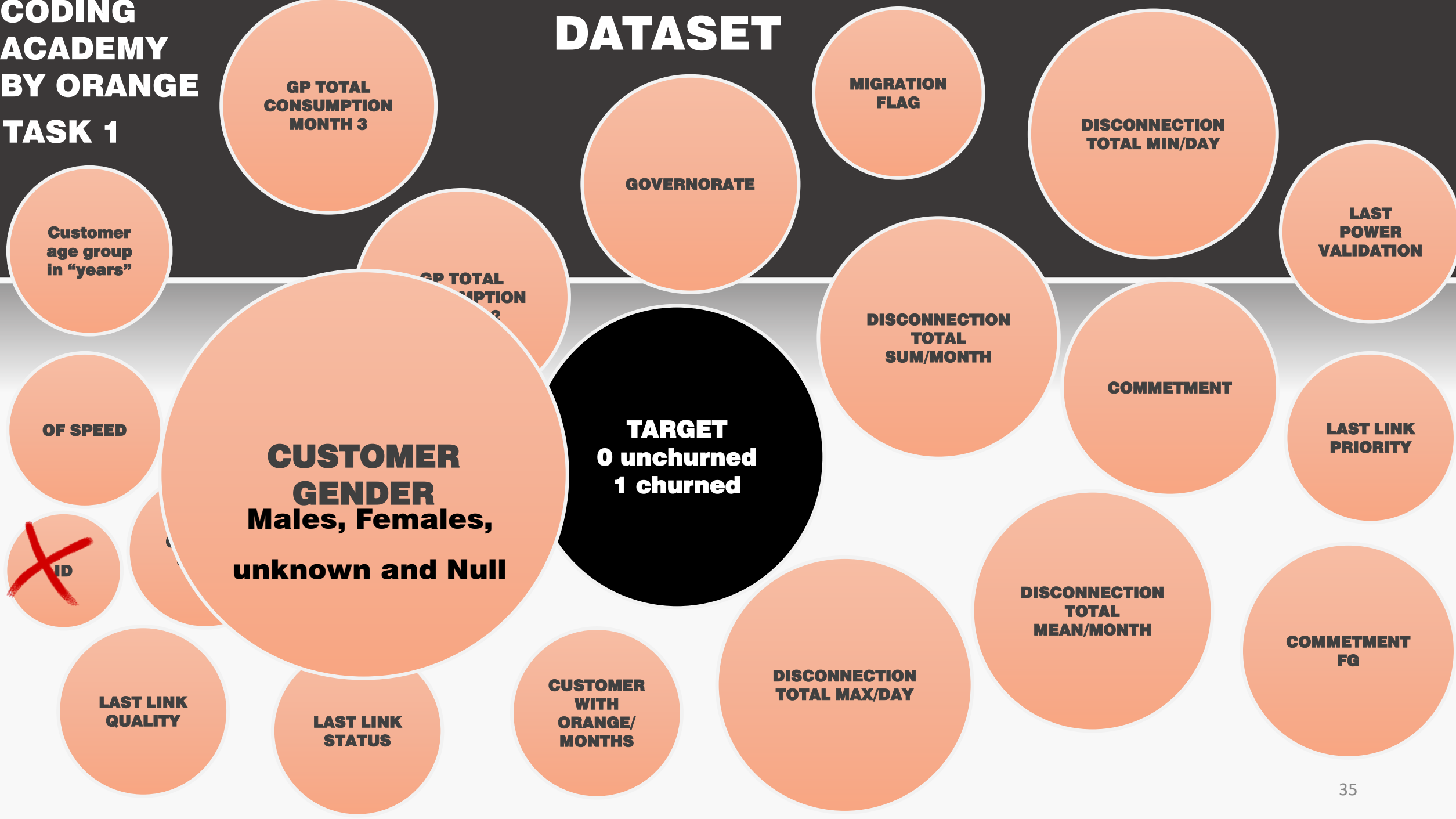
**CODING
ACADEMY
BY ORANGE
TASK 1**

DATASET



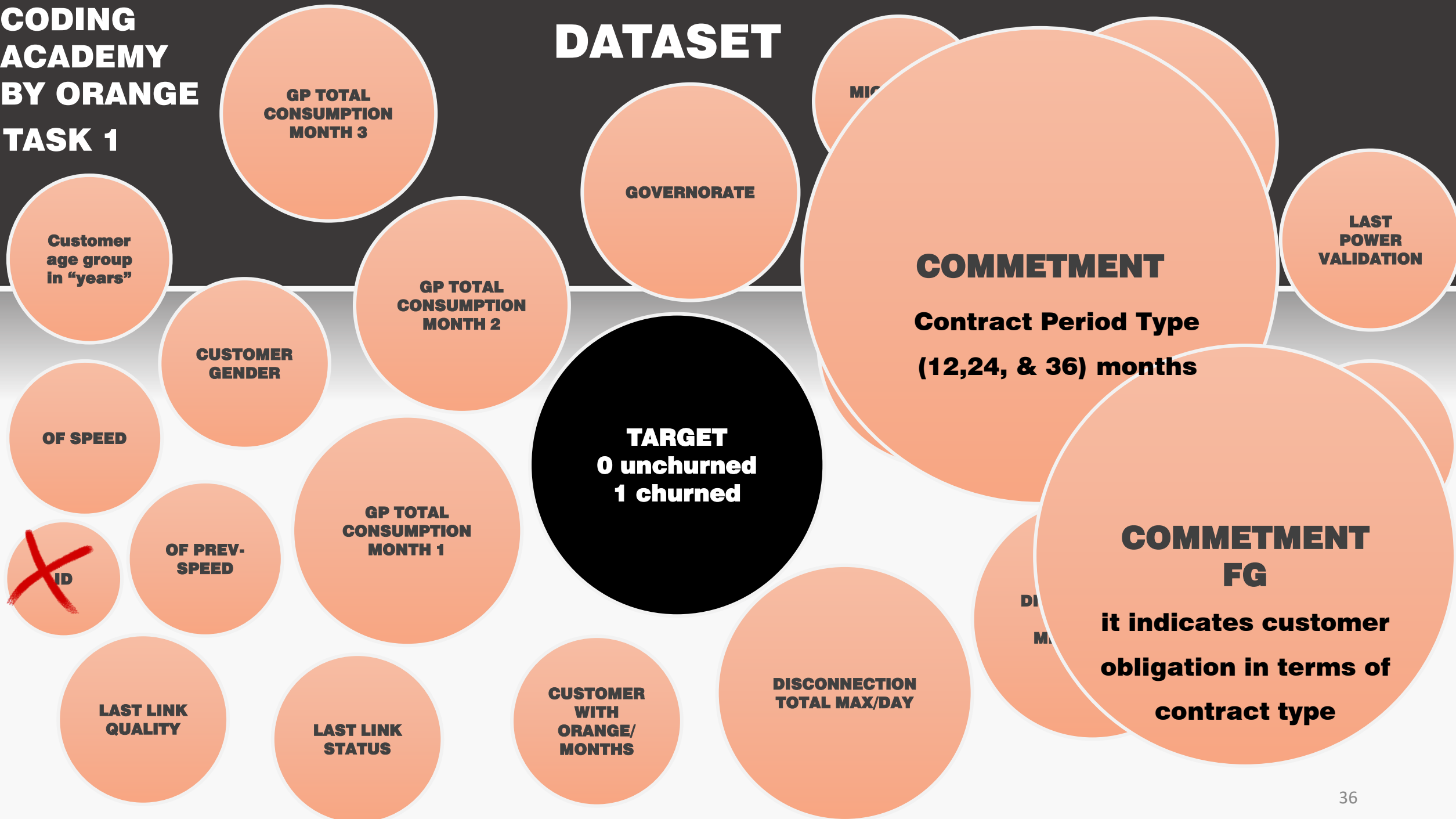
**CODING
ACADEMY
BY ORANGE
TASK 1**

DATASET

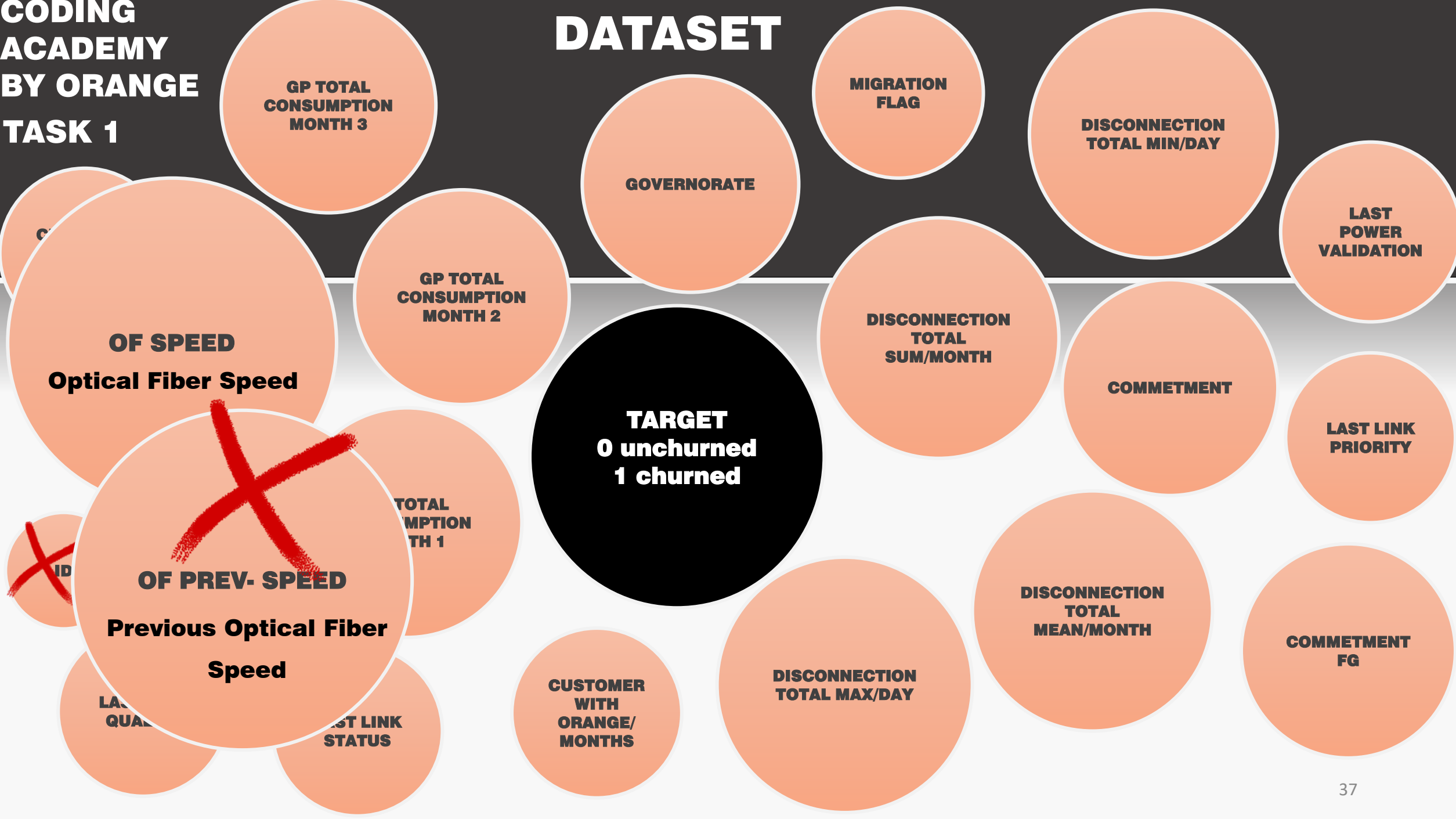


**CODING
ACADEMY
BY ORANGE
TASK 1**

DATASET

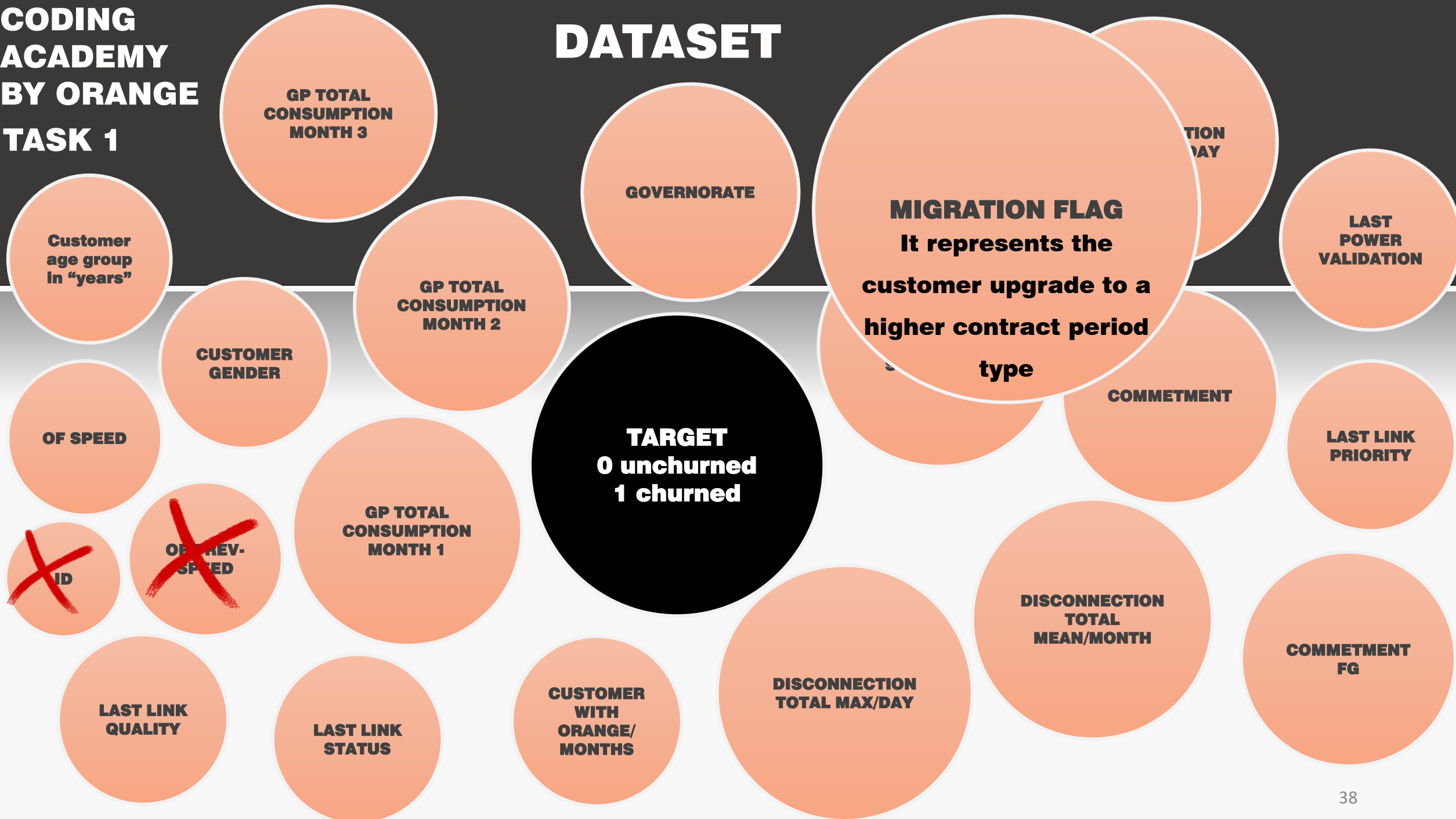


DATASET



**CODING
ACADEMY
BY ORANGE
TASK 1**

DATASET





DATASET

**GP TOTAL
CONSUMPTION
MONTH 3**

**Customer
age growth
in “years”**

GOVERNORATE

MIGRATION FLAG

**DISCONNECTION
TOTAL MIN/DAY**

BASE POWER VALIDATION

**GP TOTAL
CONSUMPTION
MONTH 2**

GET Burned Churned

**DISCONNECTION
TOTAL
SUM/MONTH**

COMMETMENT

**LAST LINK
PRIORITY**

OF SPEED

CUSTOMER GENDER

**GP TOTAL
CONSUMPTION
MONTH 1**

**DISCONNECTION
TOTAL MAX/DAY**

**DISCONNECTION
TOTAL
MEAN/MONTH**

COMMETMENT FG

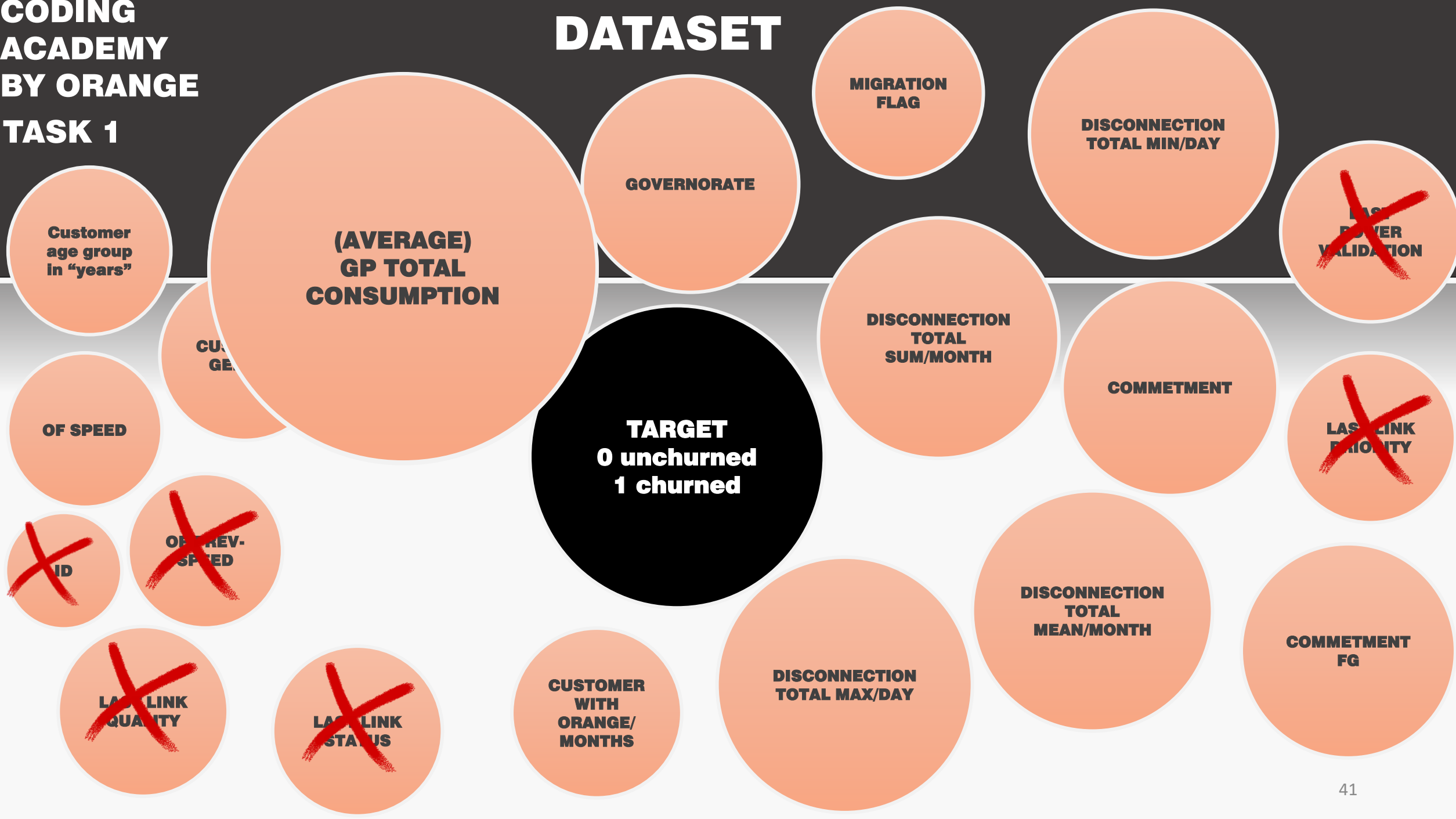
**OFF-REV-
SPEED**

ID

**LAST LINK
QUALITY**



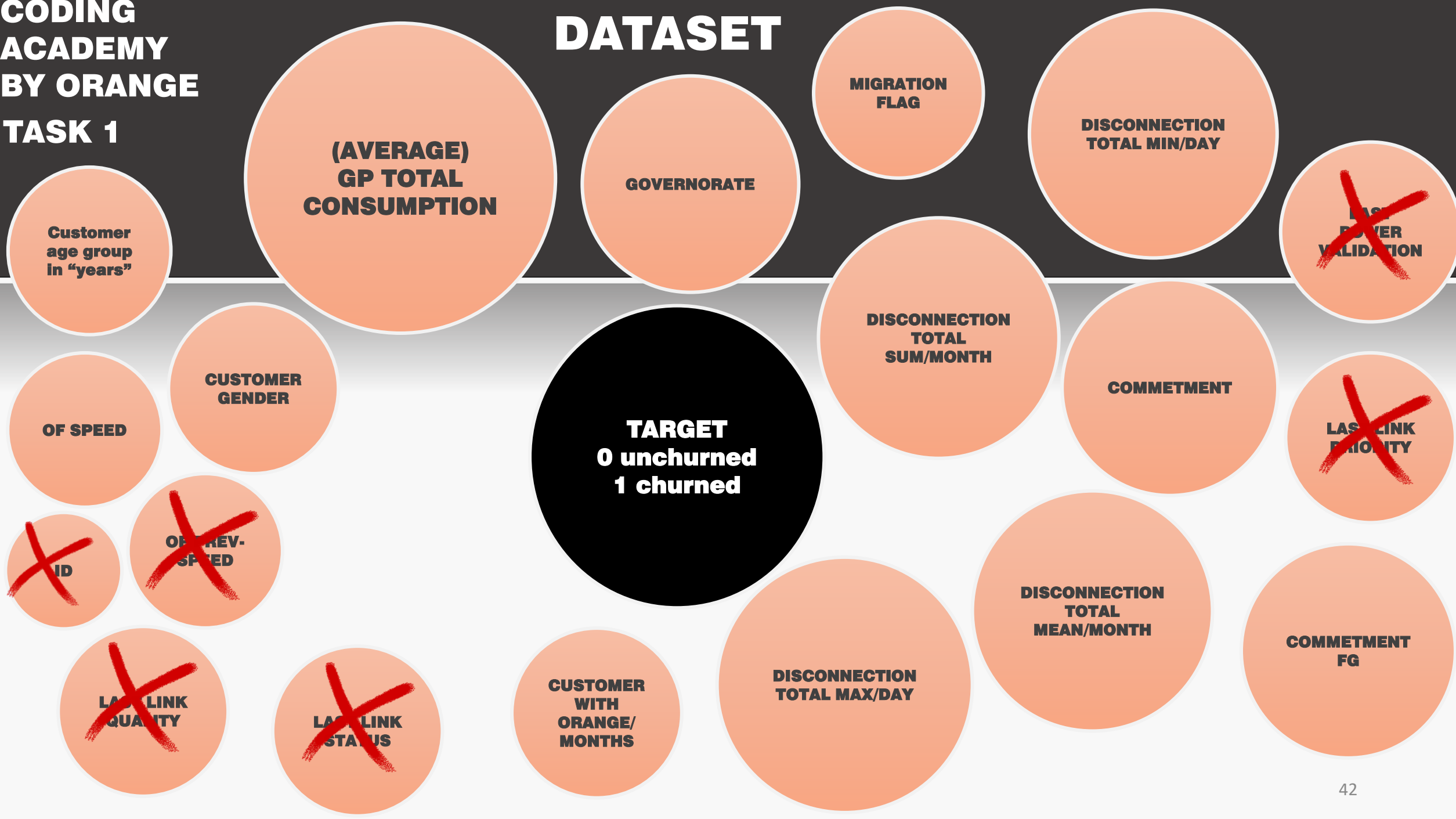
DATASET



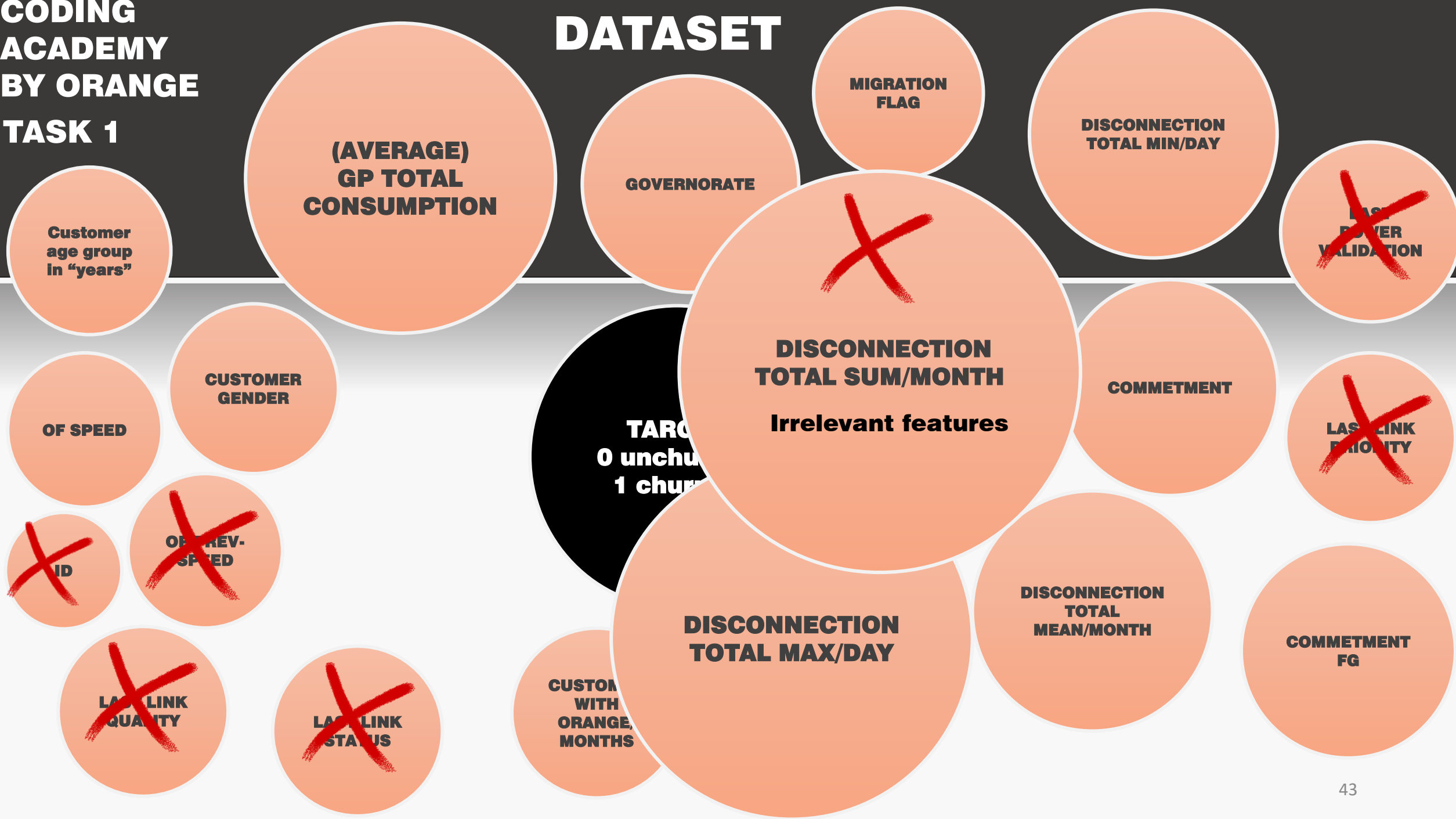
**CODING
ACADEMY
BY ORANGE**

TASK 1

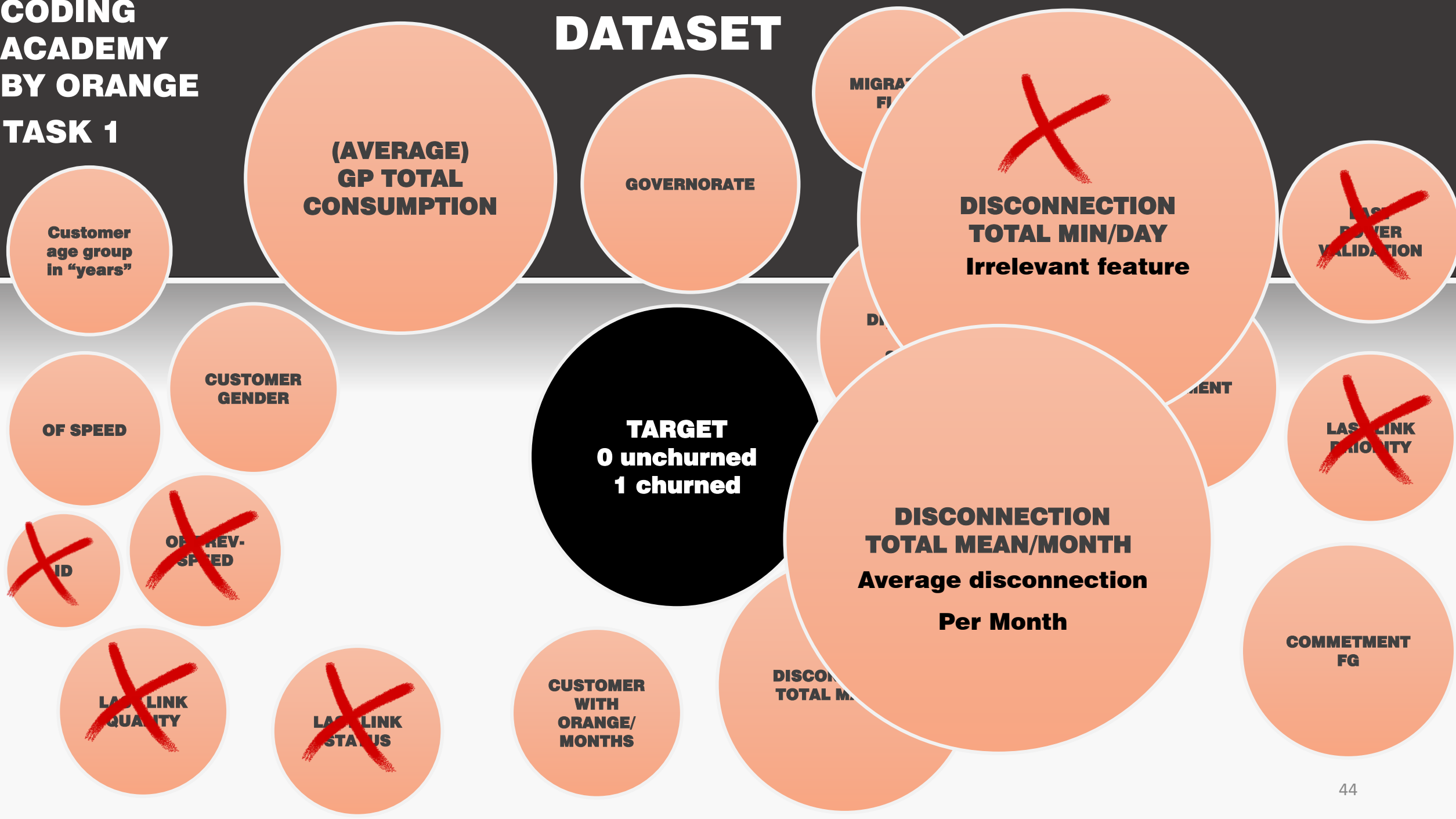
DATASET



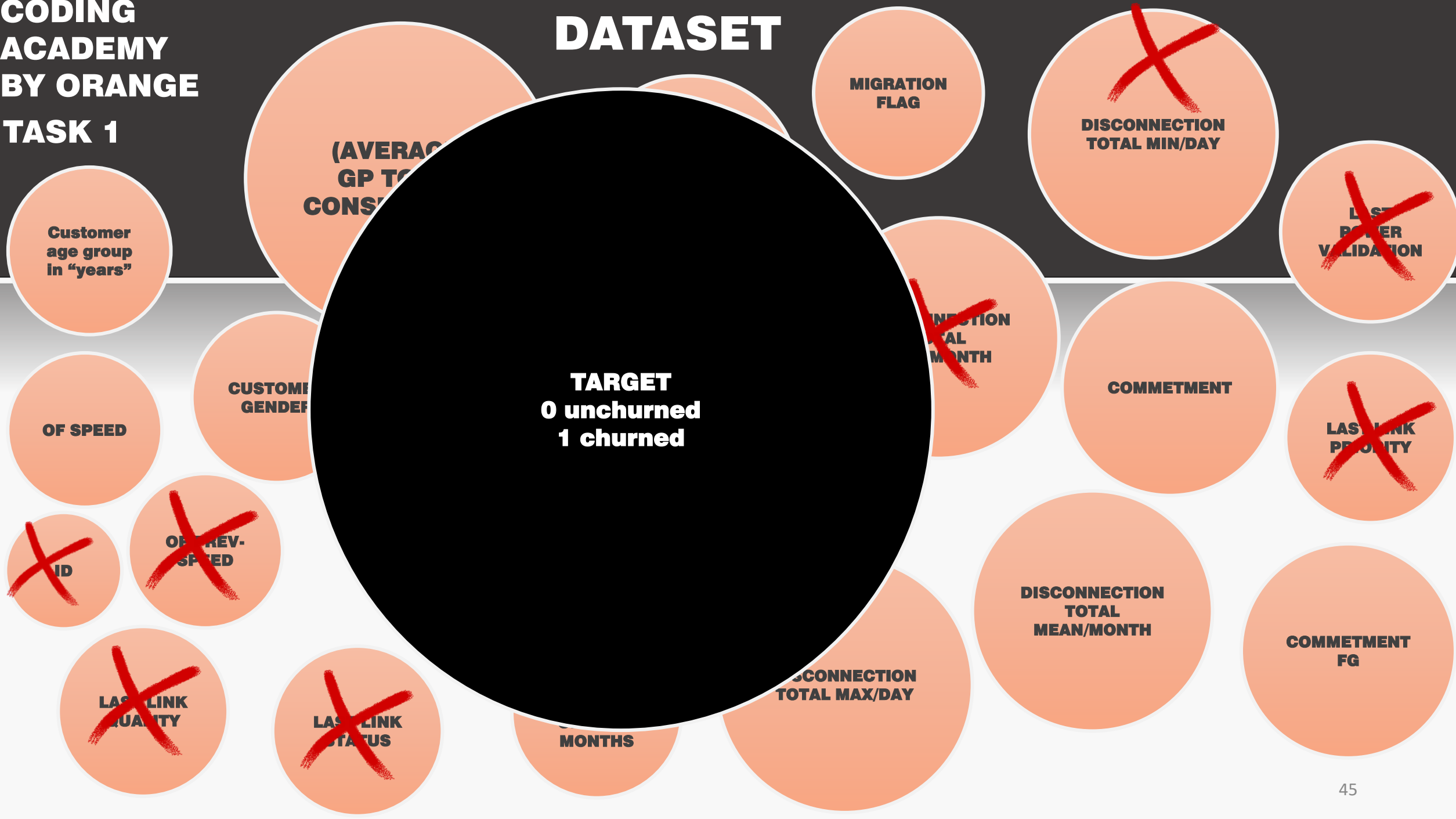
DATASET

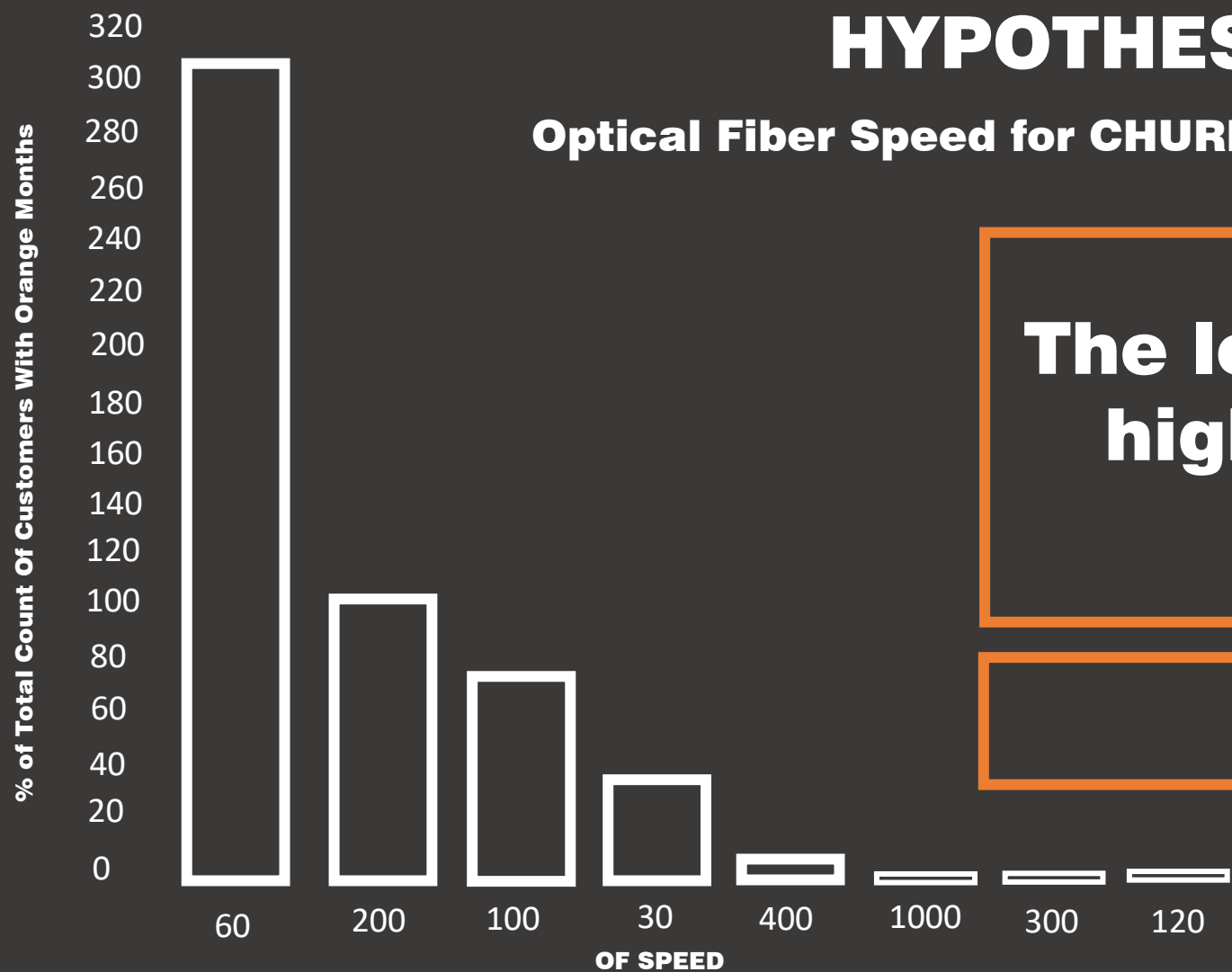
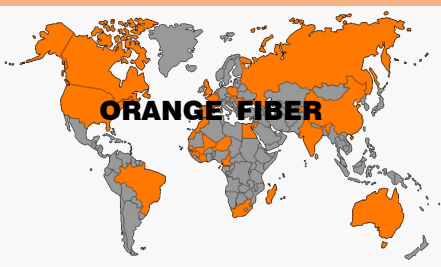


DATASET



DATASET



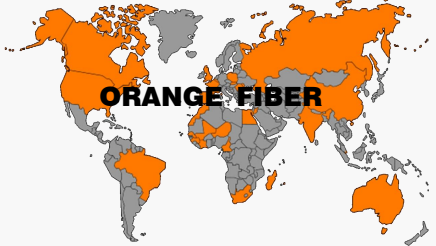


HYPOTHESIS

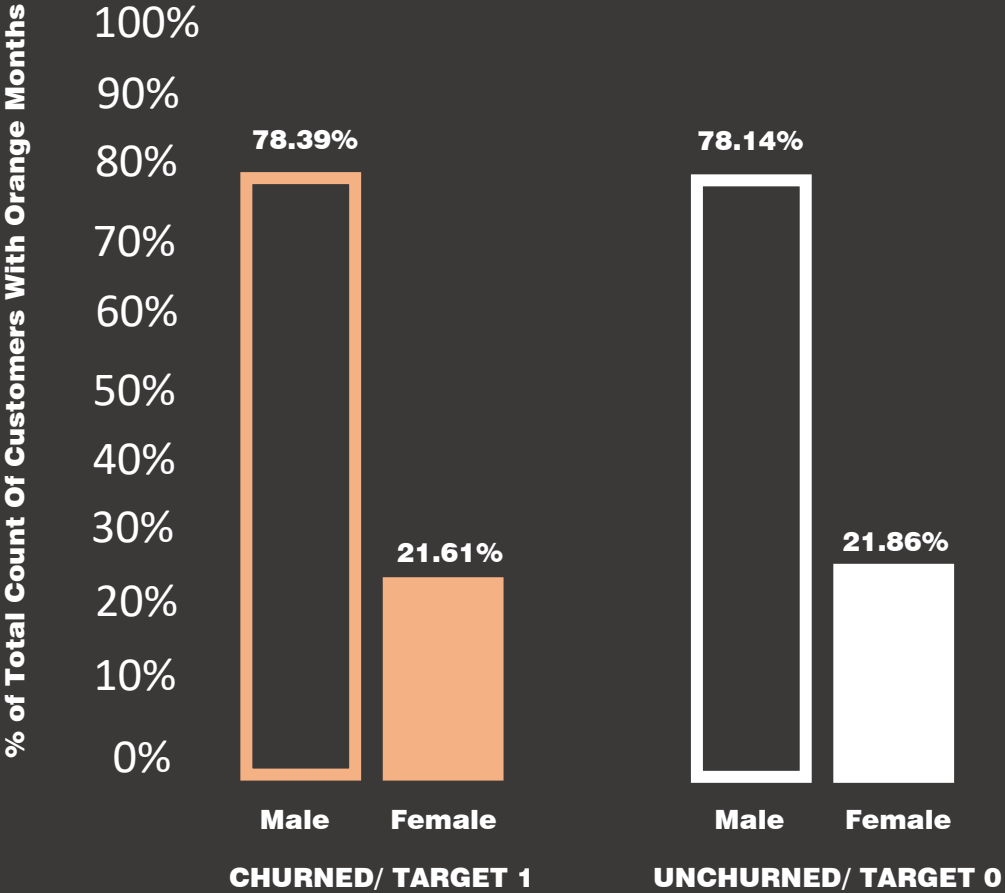
Optical Fiber Speed for CHURNED CUSTOMERS

The lower the OF-SPEED, the higher a customer being churned

CORRECT

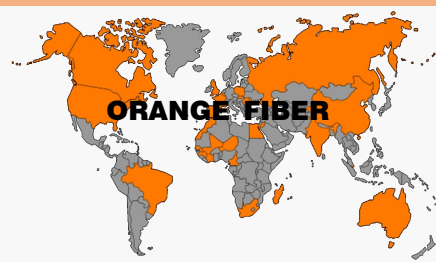


HYPOTHESIS



The probability of a male customer to be churned is higher than the probability of a female customer to be churned

WRONG

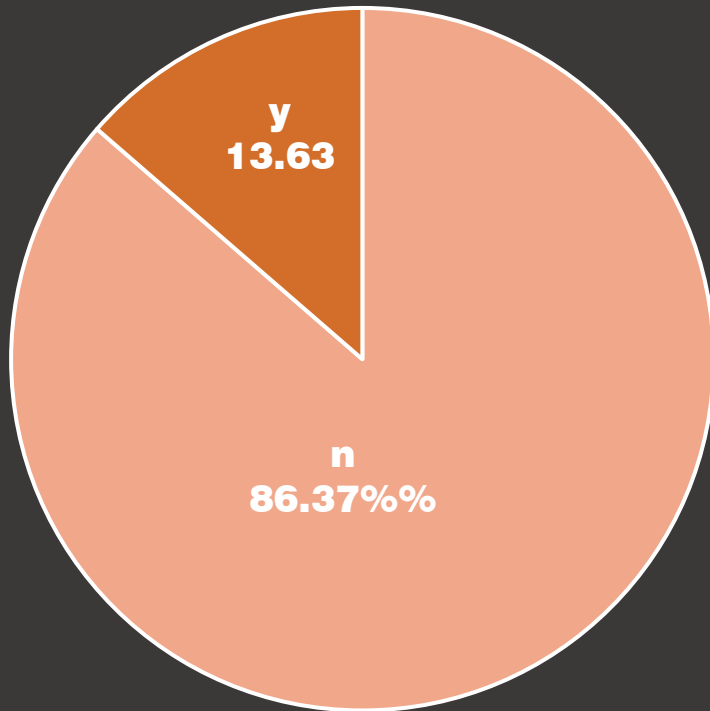


3



HYPOTHESIS

Churned Customers Who Either Upgraded Their Contract Or Not

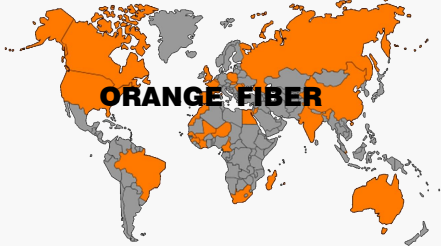


n: did not
upgrade
y: upgraded

**Customers who don't
perform migration flag (n),
have a higher probability to
being churned**

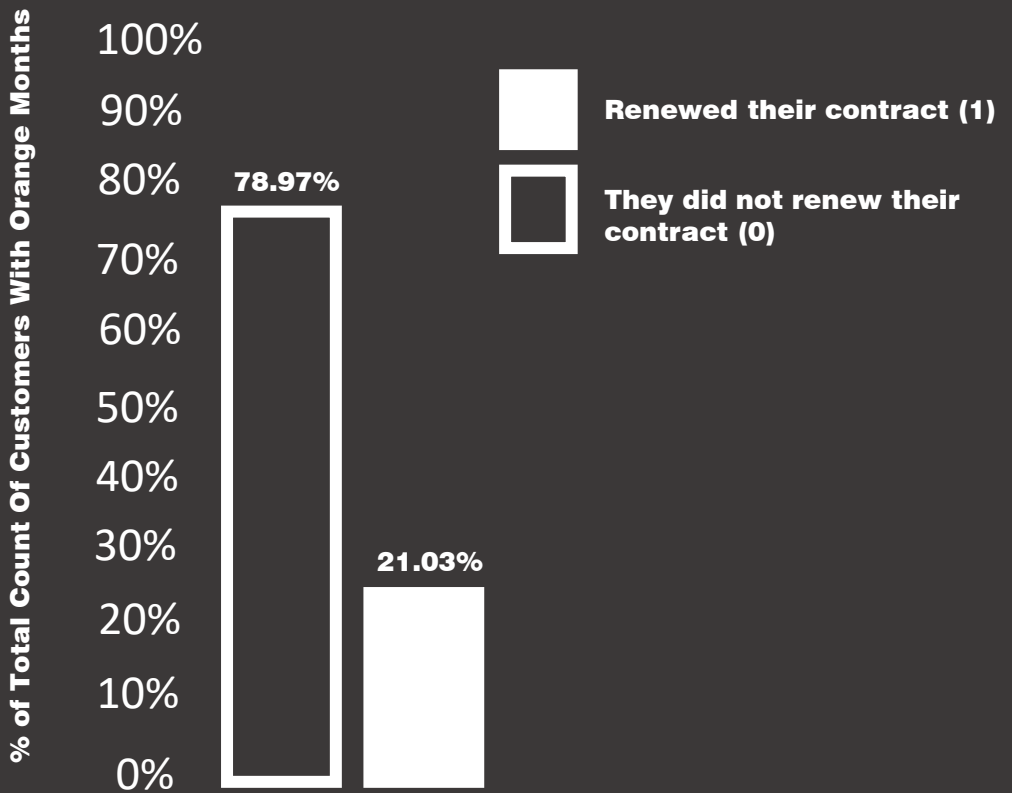
**Migration Flag: The customers who upgraded
their contract**

CORRECT



HYPOTHESIS

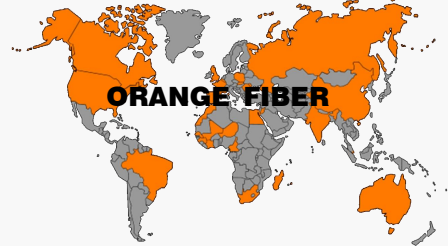
Churned Customers Who Either Renewed Their Contracts Or Not



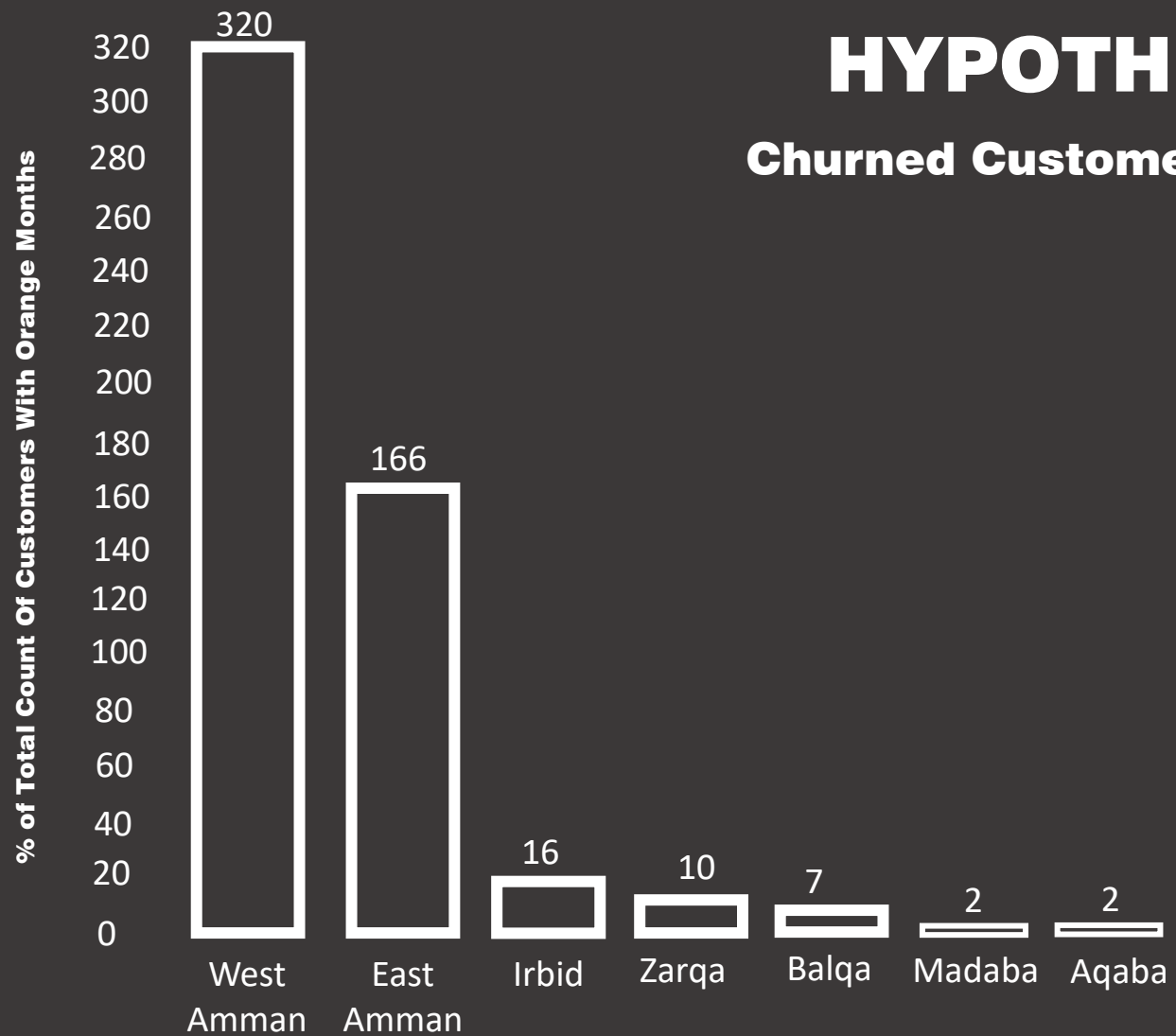
CHURNED CUSTOMERS (who left Orange)

Customers who don't commit with their contract period are most likely to being churned more than the customers who commit with their contract period

CORRECT

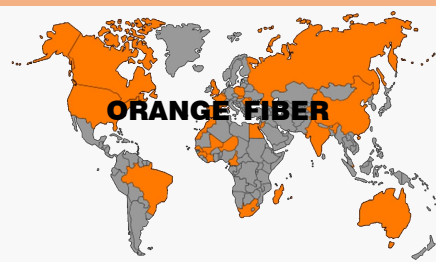


HYPOTHESIS
Churned Customers In Jordan



Higher churned customers are from West Amman

WE CAN NOT PROVE IT

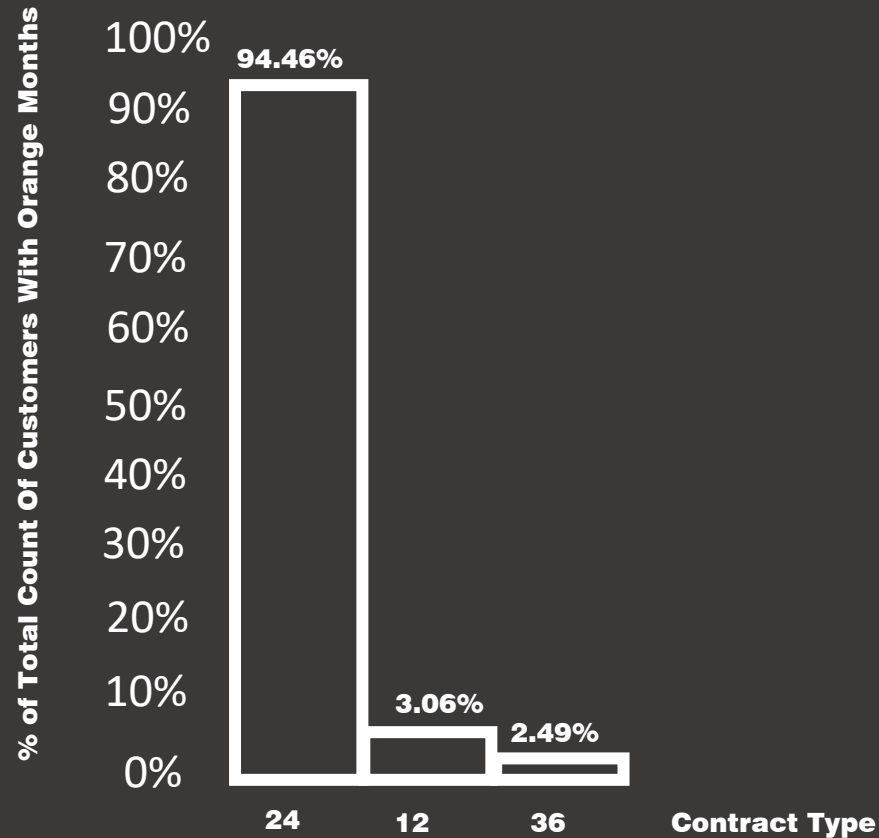


6



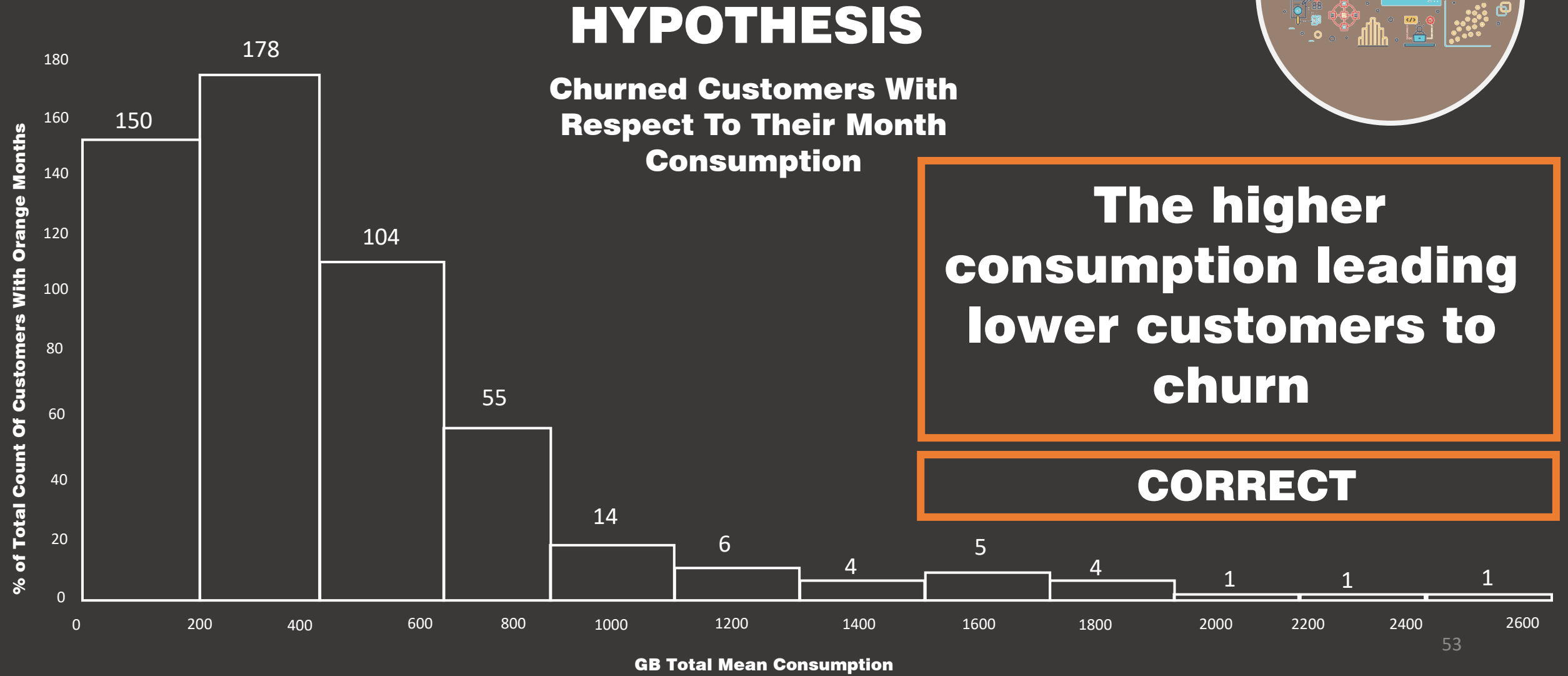
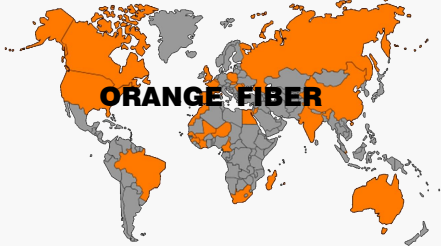
HYPOTHESIS

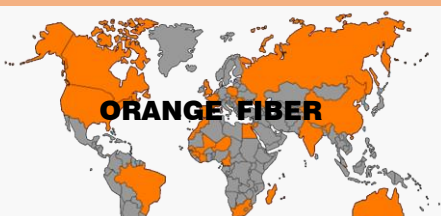
Churned Customers vs contract type



**Customers with 24
months contract
type are having
higher probability
to churn**

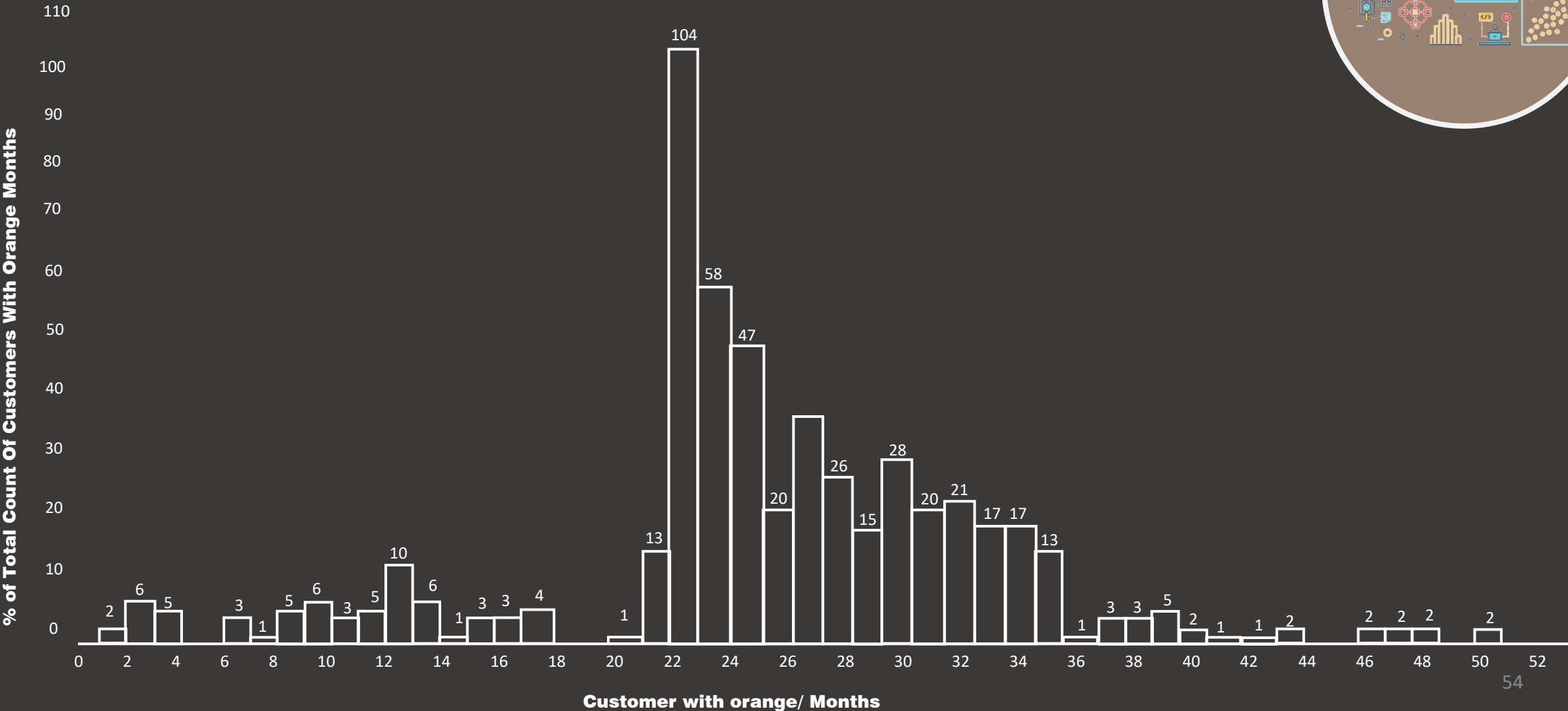
CORRECT

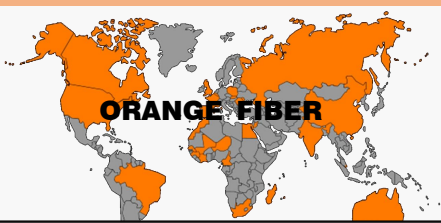




TASK 1

HYPOTHESIS



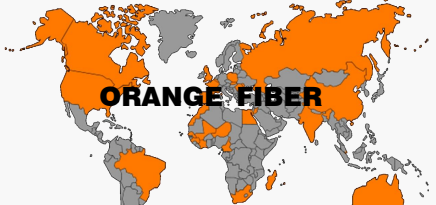


TASK 1

HYPOTHESIS

Most Customers tend to churn when passing the end of the contract period to avoid penalties and especially for contracts with 24 months because it has best offers and price



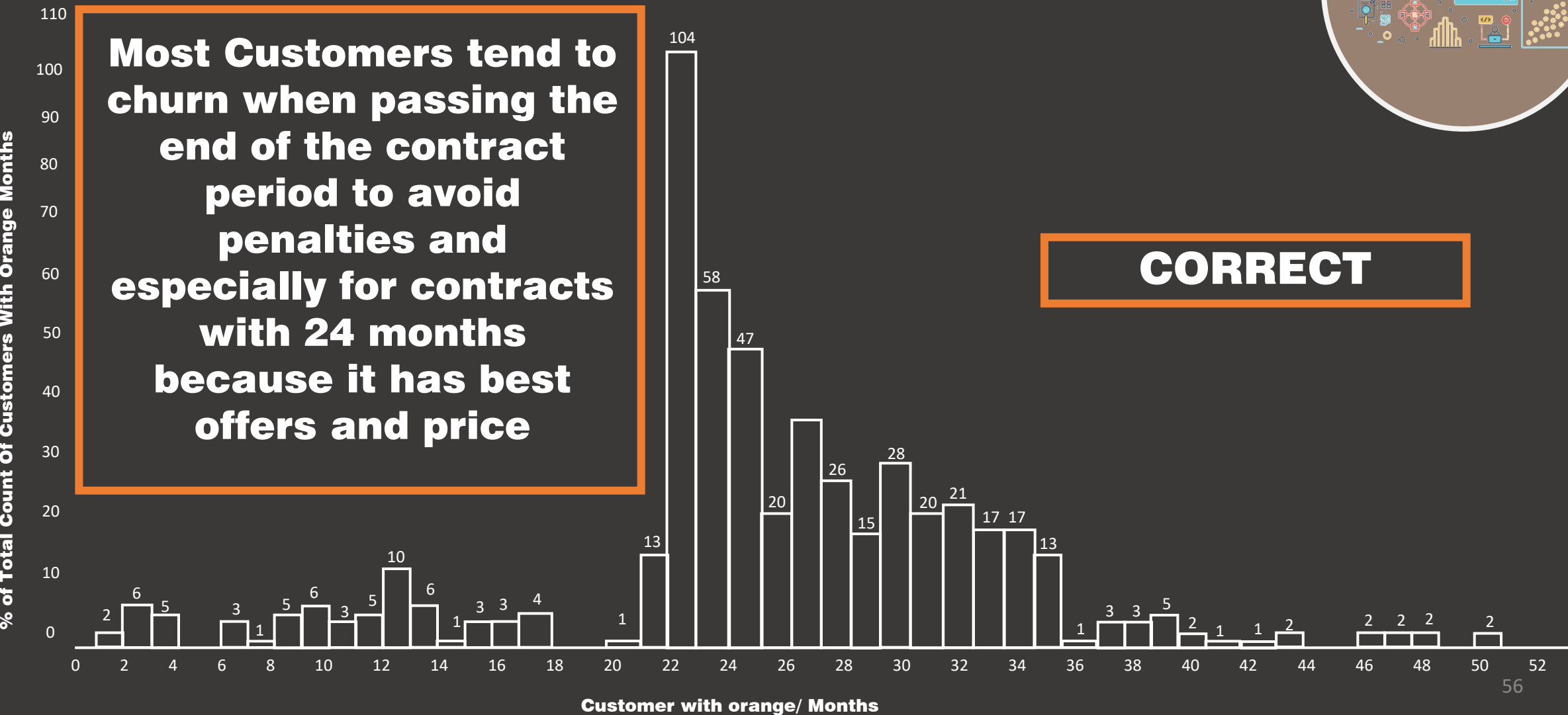


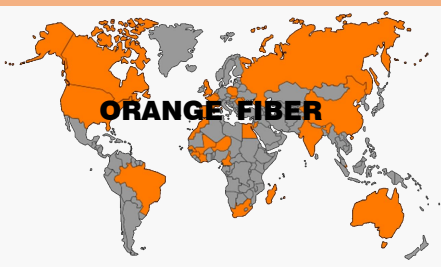
TASK 1

HYPOTHESIS

Most Customers tend to churn when passing the end of the contract period to avoid penalties and especially for contracts with 24 months because it has best offers and price

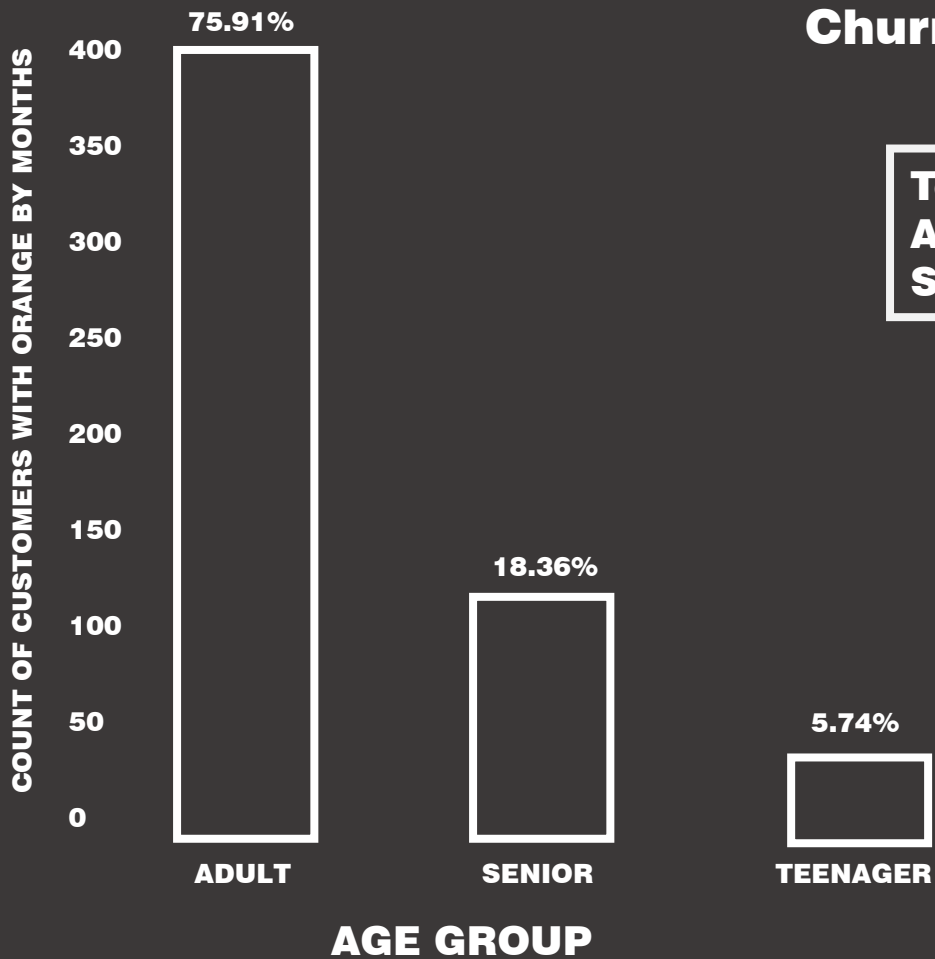
CORRECT





HYPOTHESIS

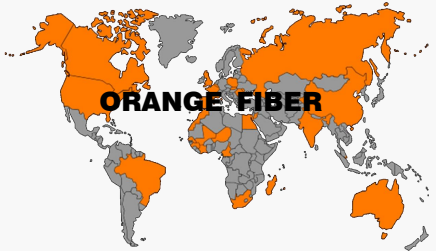
Churned Customers age group



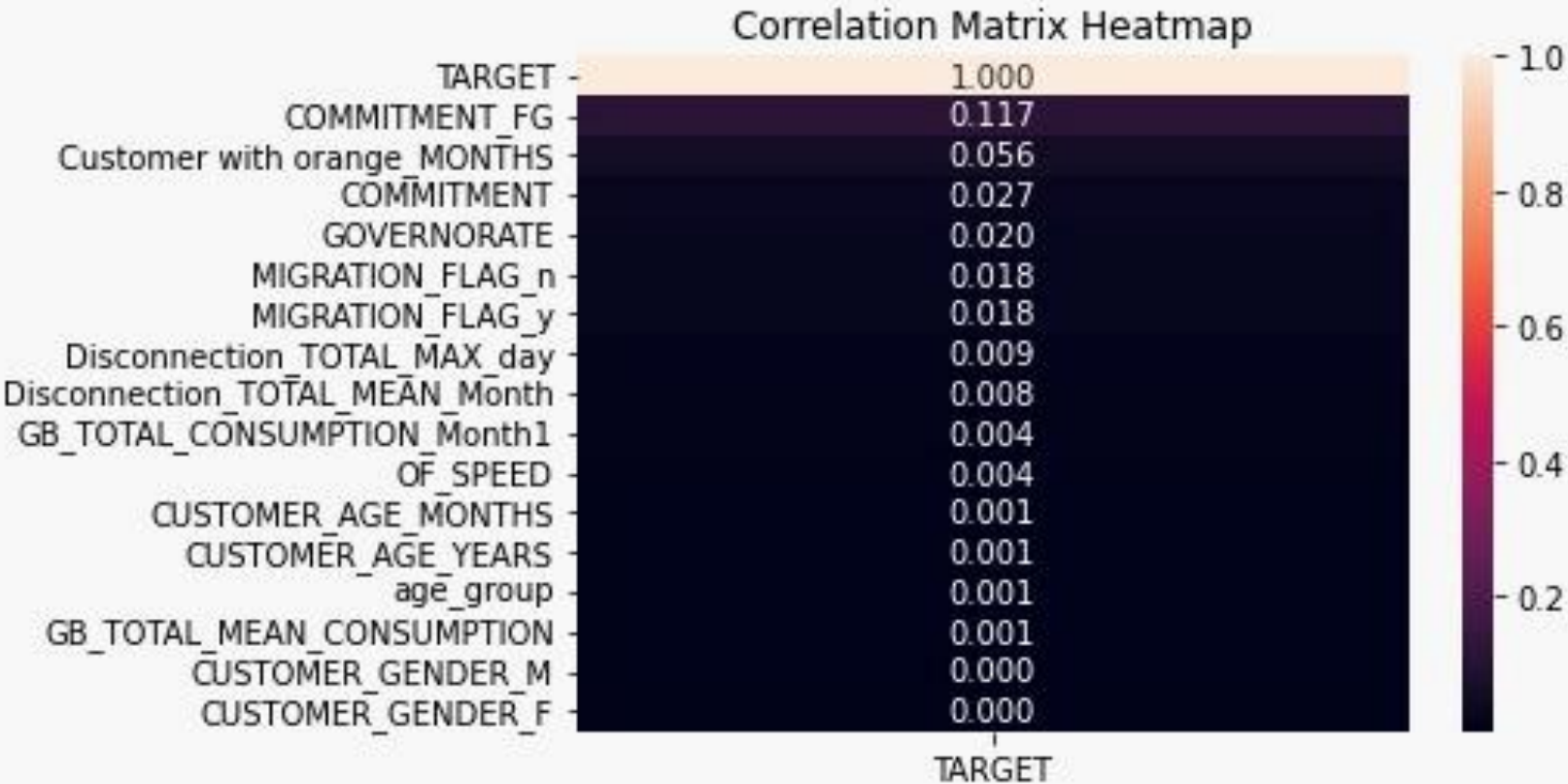
Teenager	19-25
Adult	25-60
Seniors	Above 60

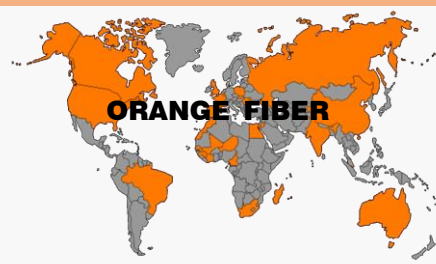
Adults are the “age group” that churned the most

CORRECT



CORRELATION MATRIX HEATMAP





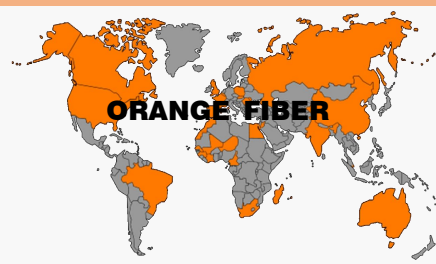
Future work

1- track down who participated in the data collection and who made the data.

2-collecting more features for our customers

- time-based features to capture the patterns and trends in customer behavior over time.**
- education level, and occupation could provide insights into which customer groups are more likely to churn.**
- customer service interactions, such as the number of complaints or inquiries, can provide insights into customer satisfaction with the service and identify areas for improvement.**
- whether a customer has a spouse or not**

3- Incorporate unstructured data such as social media or customer service transcripts that can provide insights into customer behavior and attitudes.



RECOMMENDATION

Ensuring the Supply Chain & Technical Departments are providing a high of-speed all the time, and continuous check for any technical issue could lead to disconnection, because these two are the main factors could lead a customer to churn

Offering attractive deals to customers before the end of their contract is a proactive way to prevent churn

it's important to note that churn rate is not the only metric that companies should consider when evaluating customer satisfaction and loyalty. Other factors such as customer engagement, customer lifetime value, and customer referral rates should also be taken into account

accurate measurement of churn rate requires careful consideration of the time interval and definition of churn, as well as the data quality and business context.