

# **Costumers Satisfaction Survey**

**Mireille P. Feudjio T.**

**DATA SCIENTIST/AI RESIDENT  
APZIVA**

# Outline

- **Context and Problem statement**
- **Exploratory data analysis**
- **Hypothesis testing**
- **Features Engineering & Data Preprocessing**
- **Modelling**
- **Conclusion**

## Context and Problem statement

### Context

- There were many challenges during the COVID 19 pandemic. A young and growing company specializing in logistics and delivery strives to make its customers happy.
- With a global expansion strategy, the company needs to know what makes its customers happy.
- We conducted a survey to measure the satisfaction level of each customer. This is a crucial element in improving our operations at all levels. If we can predict what makes our customers happy or unhappy, then we can take action.

### Goals

- Predict if a customer is happy or not based on the answers they give to questions asked.
- finding which questions/features are more important when predicting a customer's happiness

### Success Metrics:

Reach 73% accuracy score or above, or convince us why your solution is superior.

We are definitely interested in every solution and insight you can provide us.

**Time** : Try to submit your working solution as soon as possible. The sooner the better.

### Scope of solution space:

model development should take into account all the features

### Constraints:

Getting feedback from customers is not easy .

# Dataset informations

## Attribute Information

data shape is (126, 7)

Y = target attribute (Y)

with values 0: unhappy customers

values 1: happy customers

X1 = my order was delivered on time

X2 = contents of my order was as I expected

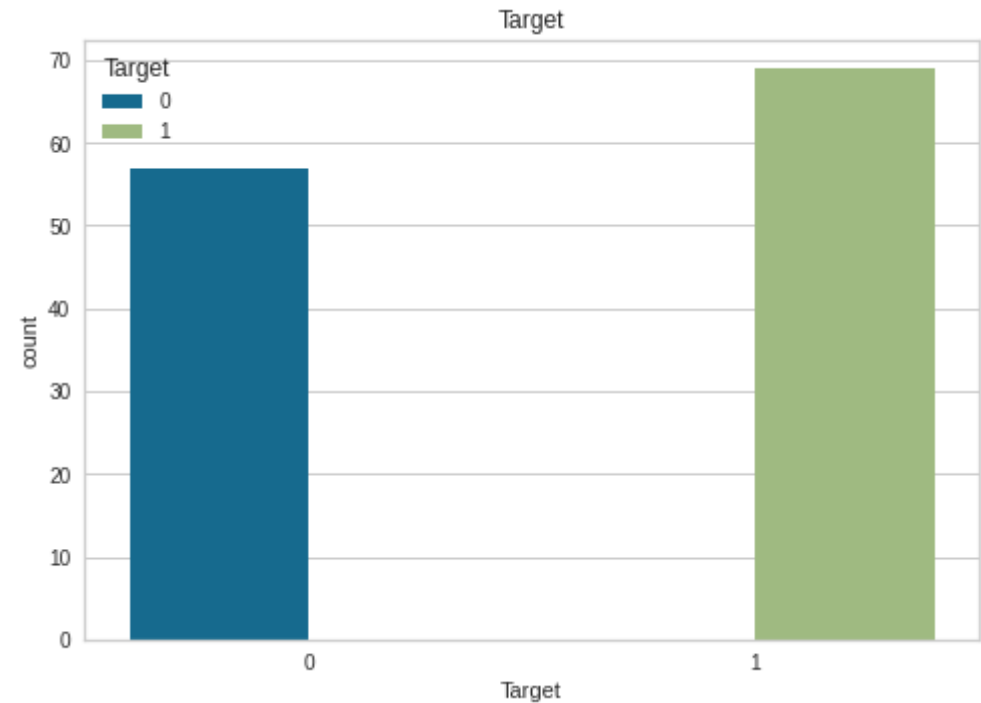
X3 = I ordered everything I wanted to order

X4 = I paid a good price for my order

X5 = I am satisfied with my courier

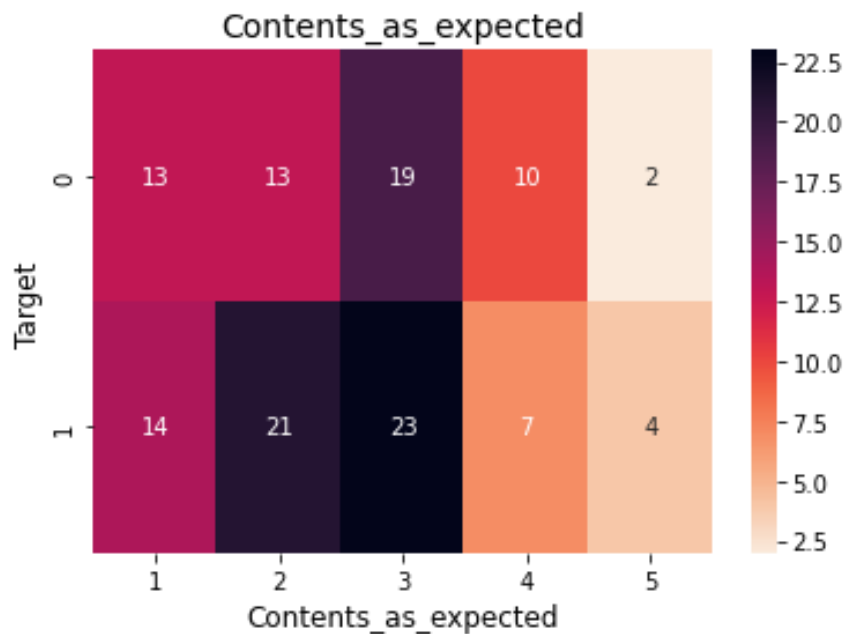
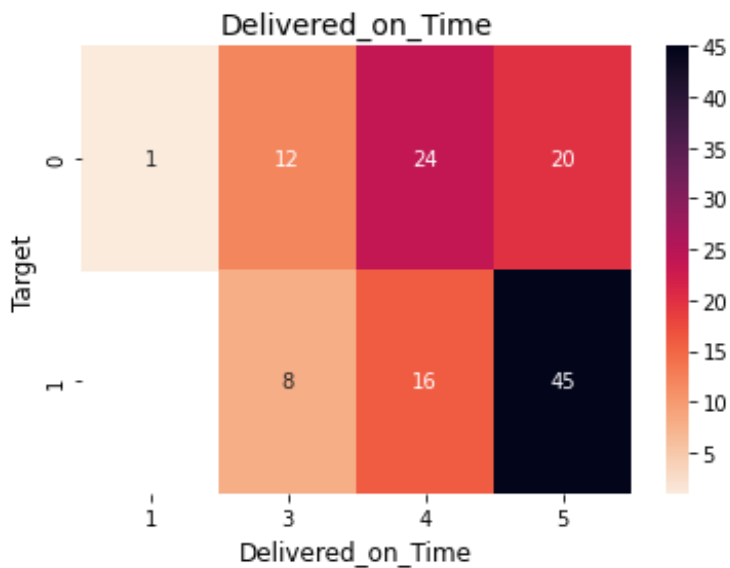
X6 = the app makes ordering easy for me

Attributes X1 to X6 indicate the responses for each question and have values from 1 to 5 where the smaller number indicates less, and the higher number indicates more towards the answer



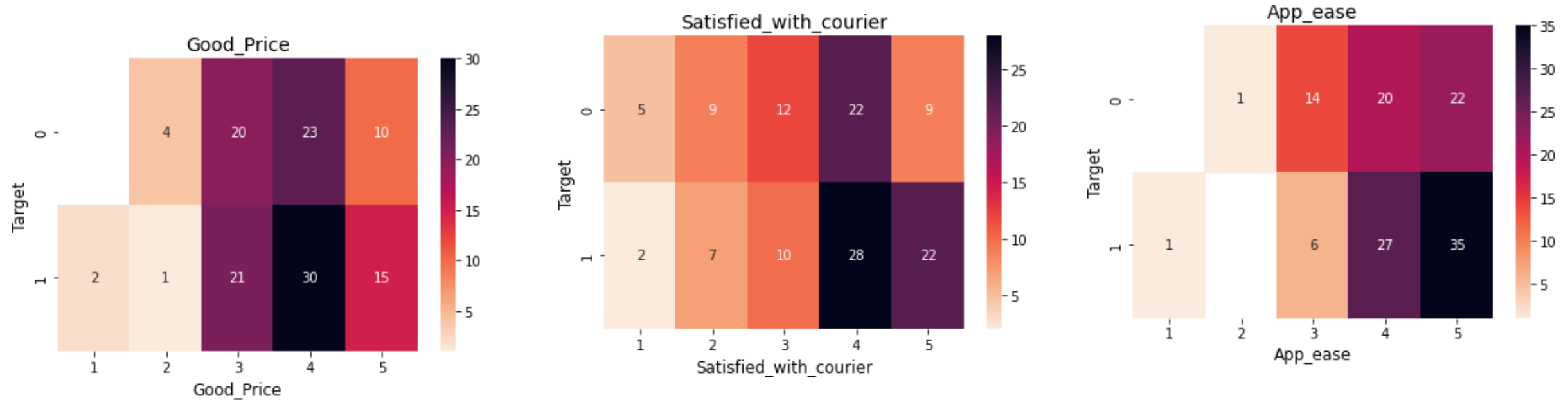
# Exploratory data analysis

What are the factors that characterize happy and unhappy customers?



# Exploratory data analysis

What are the factors that characterize happy and unhappy customers?

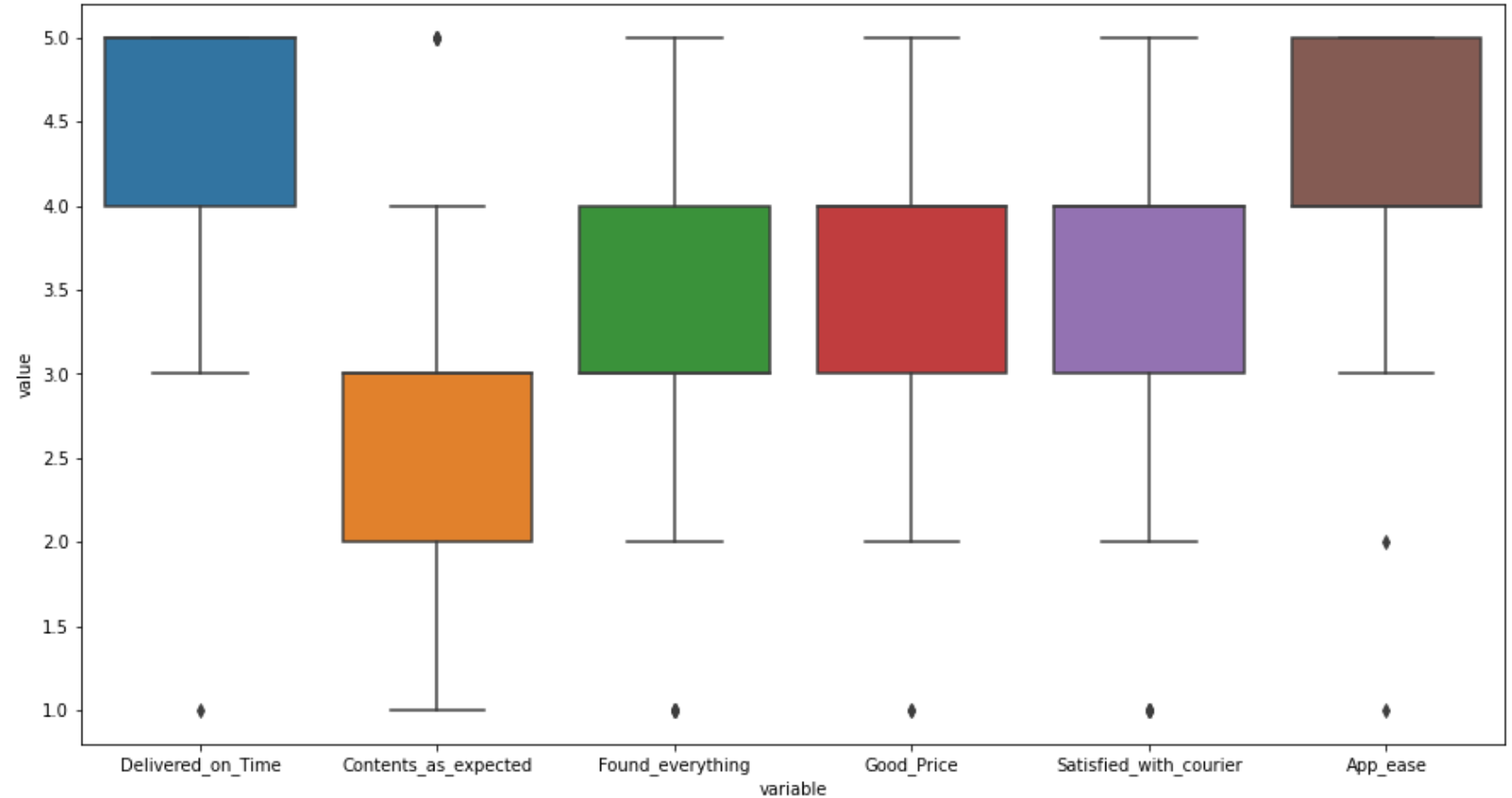


Keynote:  
variables could be considered as discrete or ordinal data, ranging from 1 to 5.

highest frequency of most features were between 3 and 5  
Exception of the variable "the content of the order was expected" which had a high frequency value between 1 and 3.

# Exploratory data analysis

Do we have outliers in the dataset?

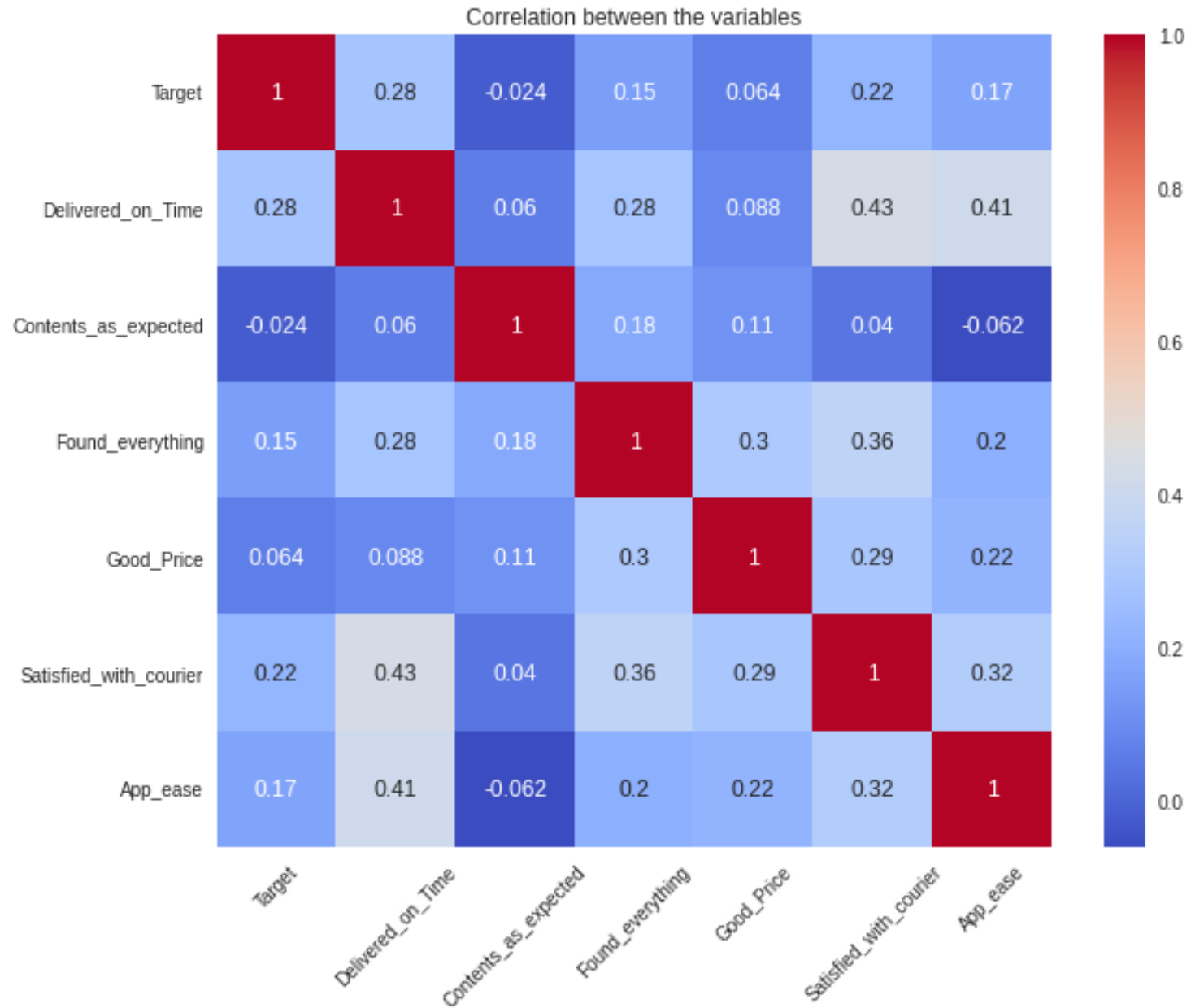


Keynote:

- variables ranging from 1 to 5
- pay attention to "Content\_as\_expected" feature due to their low score.

# Exploratory data analysis

How is the correlation between Variables ?





# Hypothesis testing : Chi-square test

Define null hypothesis

$H_0$  (Null hypothesis) – Happy/unhappy customers are independent of other variables.

$H_1$  (Alternative hypothesis) - Happy/unhappy customers are dependent on other variables.

State the significance level:  $\alpha = 0.05$

Compute the Contingency table

compute the P Value

Draw conclusion

## Conclusion of the chi\_squared test

delivered\_on\_time:  $H_0$  rejected

Contents\_as\_expected:  $H_0$  accepted

Found\_everything':  $H_0$  accepted

Good\_Price:  $H_0$  accepted

Satisfied\_with\_courier:  $H_0$  accepted

App\_ease :  $H_0$  accepted

The null hypothesis was accepted in all features except one (delivered\_on\_time) where the null hypothesis was rejected.

# Features Engineering & Data Preprocessing

## Features Engineering

- Get the mean of all scores
- Get the sum of all scores
- Get the Percentage of the overall scores

## SMOTE for Imbalanced dataset

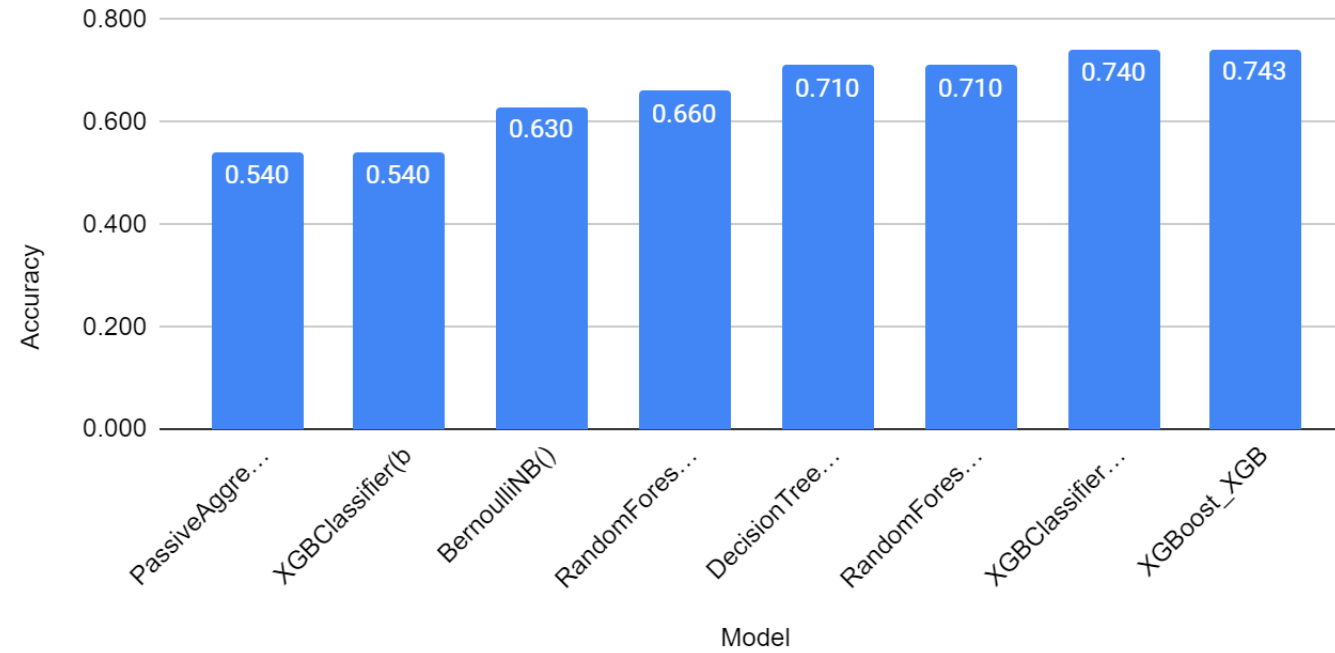
- balance the minority class
- Minority Over-Sampling
- 69 vs 69 observations

## Data Preprocessing

- Get Dummies features for ordinal variables
- Obtain target and predictors
- Divide the data into training and test sets

# Modelling

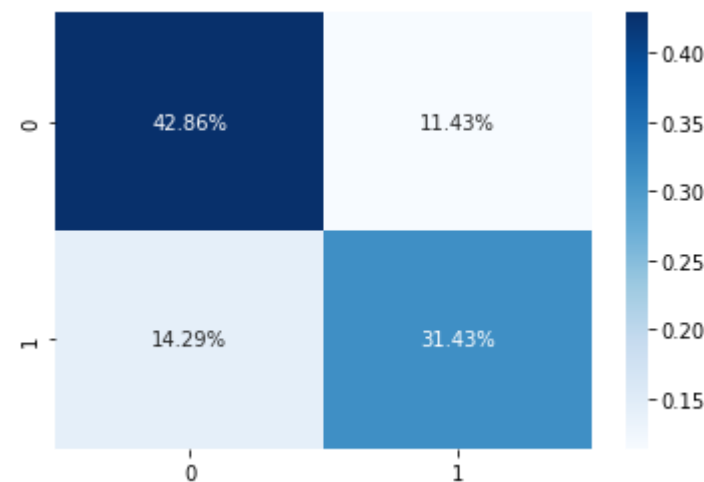
Accuracy vs. Model



# Modelling

XGBoost model

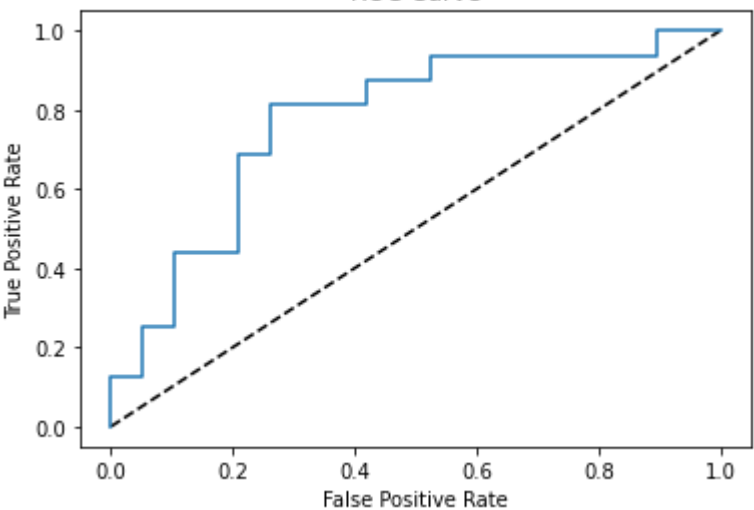
Confusion Matrix



Classification\_report with test set and predicted value

	precision	recall	f1-score	support
No_Happy	0.75	0.79	0.77	19
happy	0.73	0.69	0.71	16
accuracy	0.74			35
macro avg	0.74	0.74	0.74	35
weighted avg	0.74	0.74	0.74	35

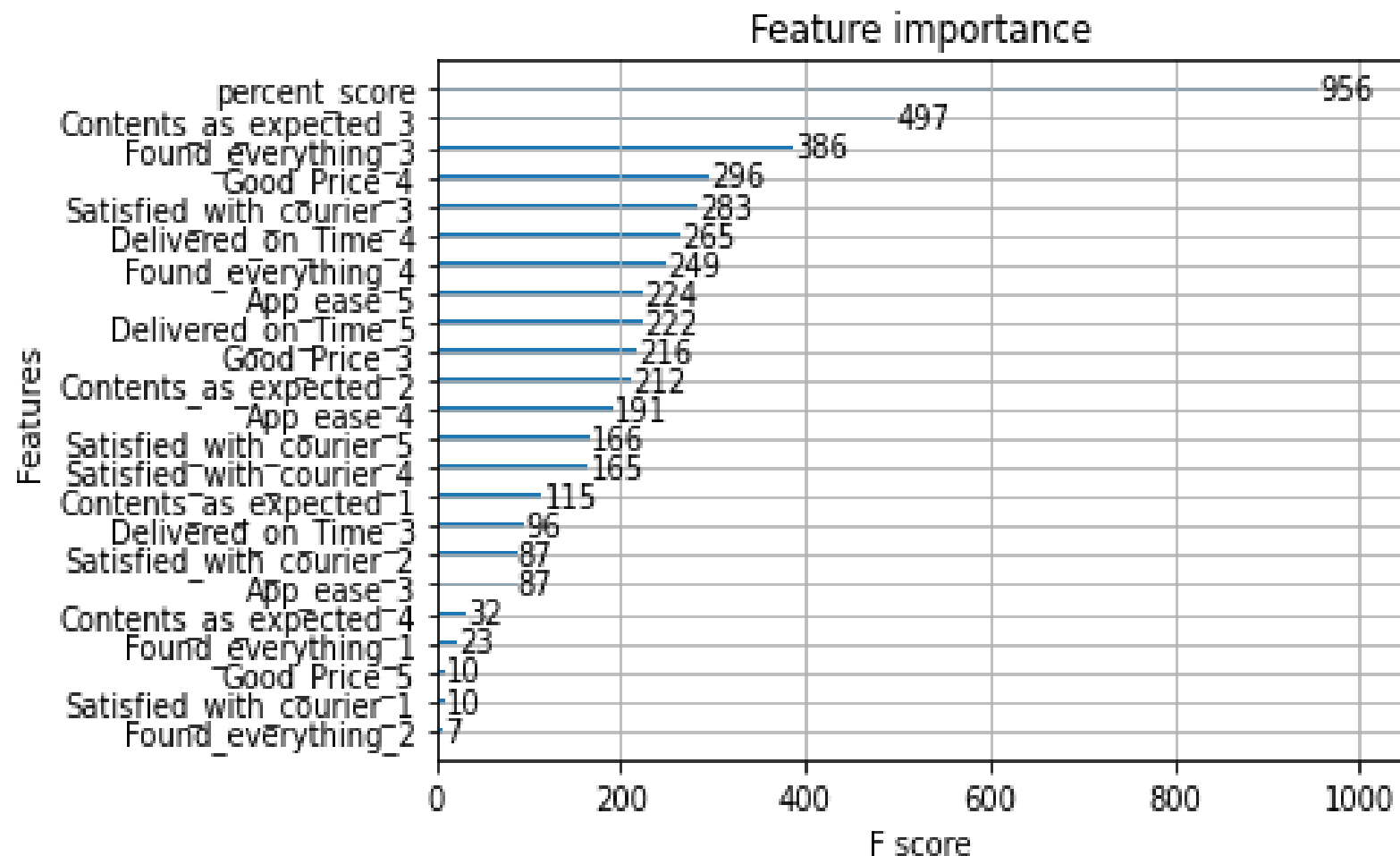
ROC Curve



AUC: 0.773

# Features of importance

XGBoost model



# Conclusion

THANKS