

CTR of an Email Campaign Predictor

Vasu Devan





**Problem
Overview**

EDA on Data

**Pre-processing
Data**

**MODEL
Selection**

**MODEL
Evaluation**

Recommendation

Overview of the Presentation

Problem Overview

- What is the Business problem?
- To Predict Click Rate of the Email
- Using insights in the business domain



Objective

- **Preforming EDA**
- **Identifying the Key Variables influencing the Problem**
- **Selecting the best Fit Model**
- **Suggesting ways to Improve the CTR**

Solution Process

The Data set from email campaign that has Details of Emails attributes like subject and body length, no. of CTA ,etc.
By Following the Steps we would provide the Click rate of the email.



Analyse Data

Exploratory Data Analysis, Label & One-Hot Encoding, Missing Values Treatment



Analytical approach

Scaling the Data, Selecting Model: Random Forest, XG Booster, Ada Booster



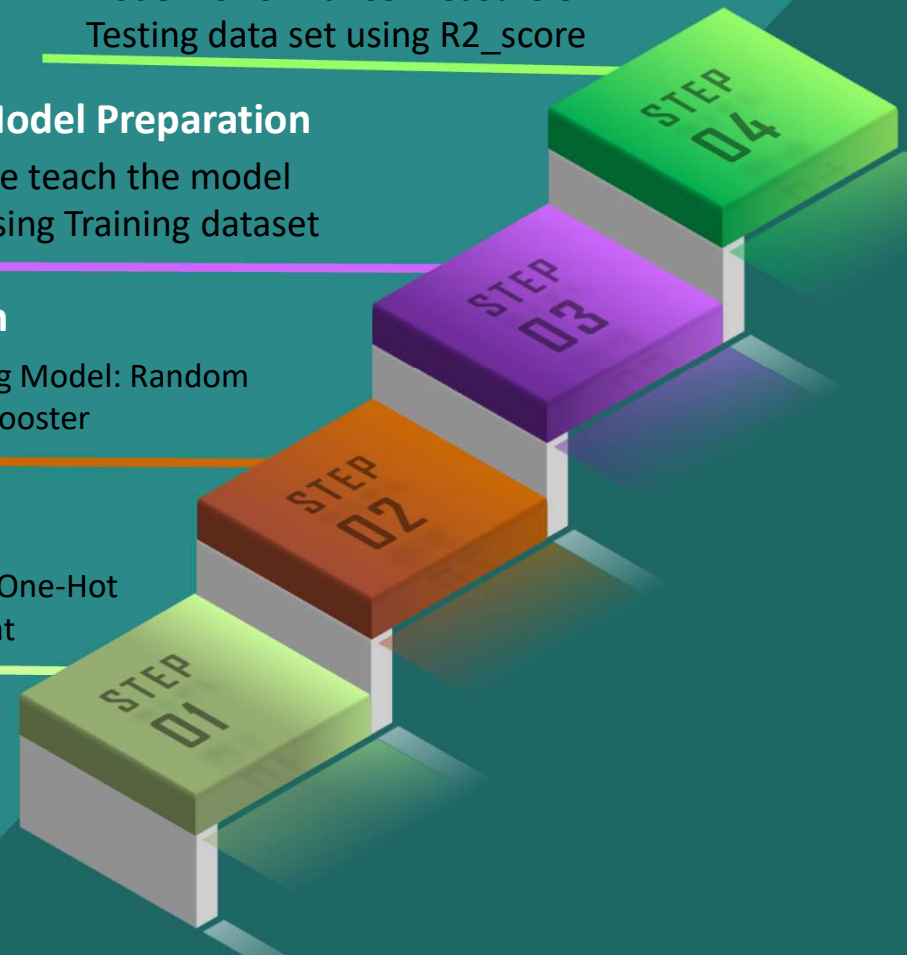
Model Preparation

We teach the model
Using Training dataset

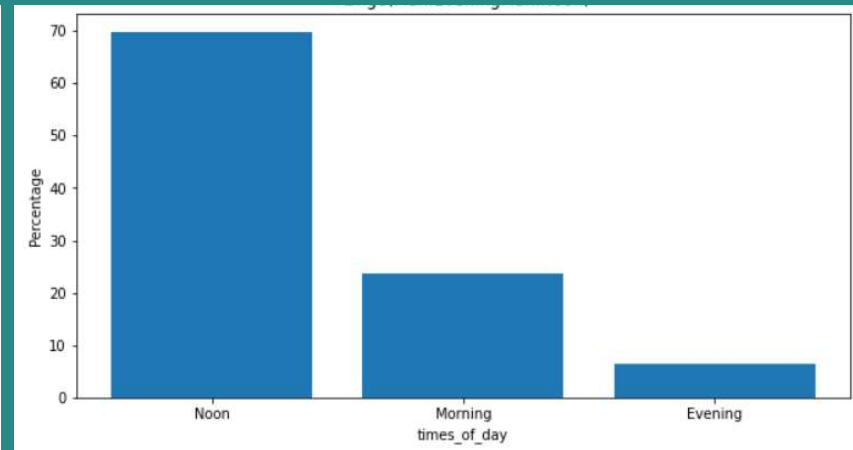
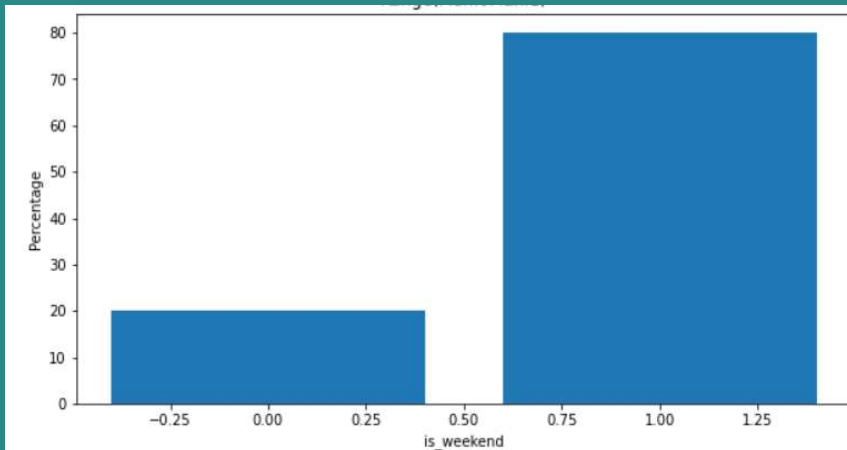
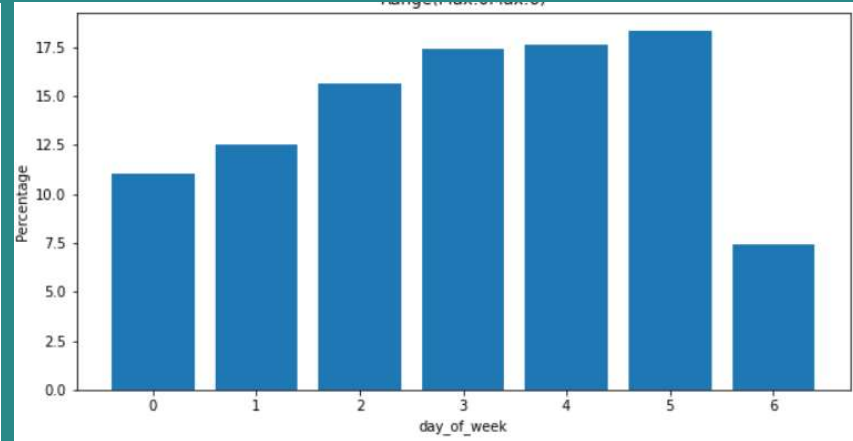
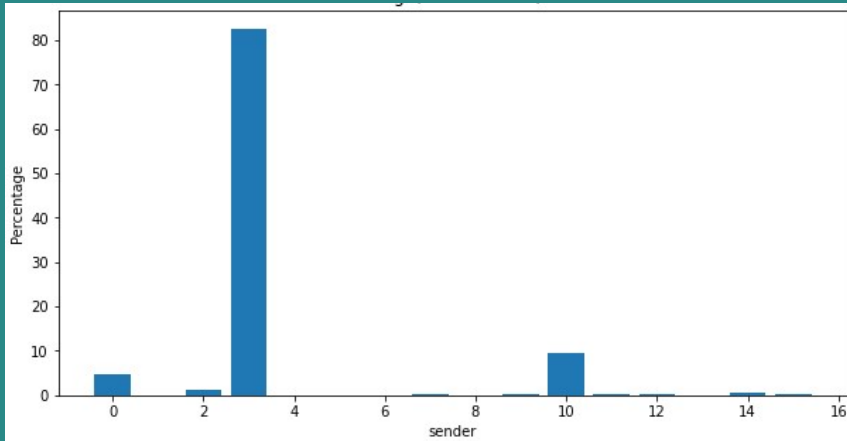


Model Validation

Model Performance measure on
Testing data set using R2_score

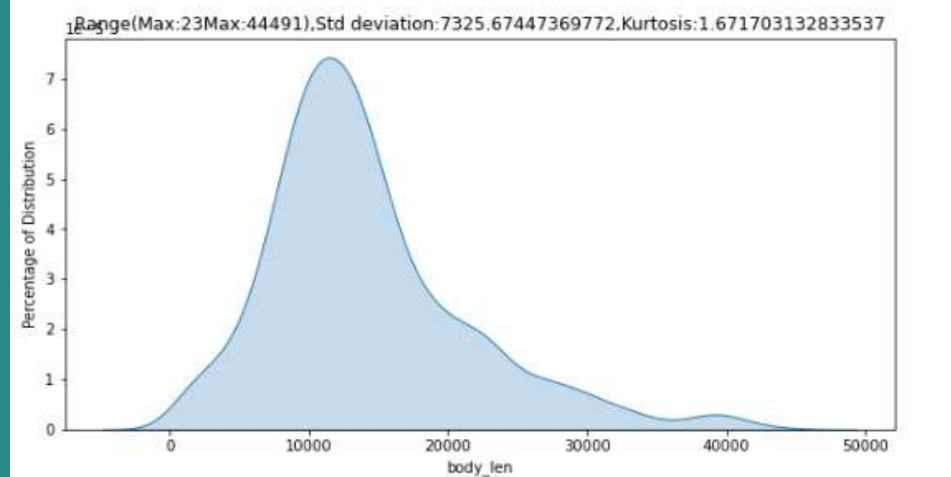
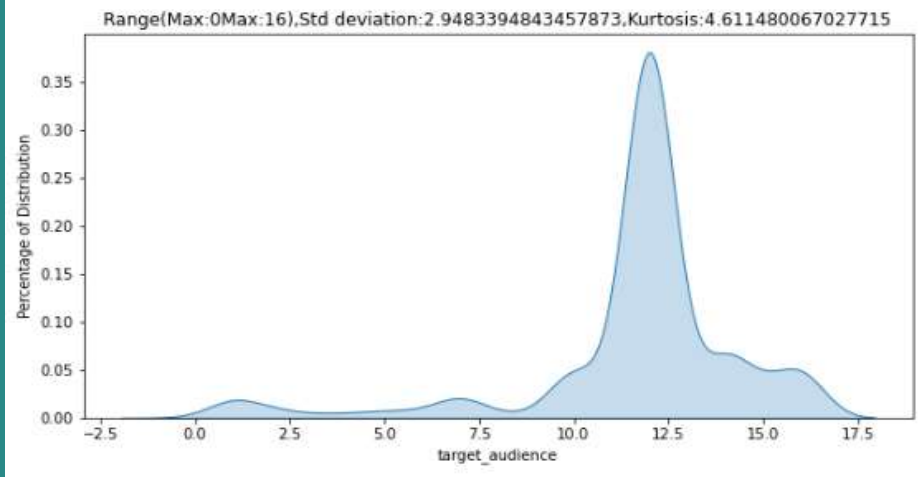
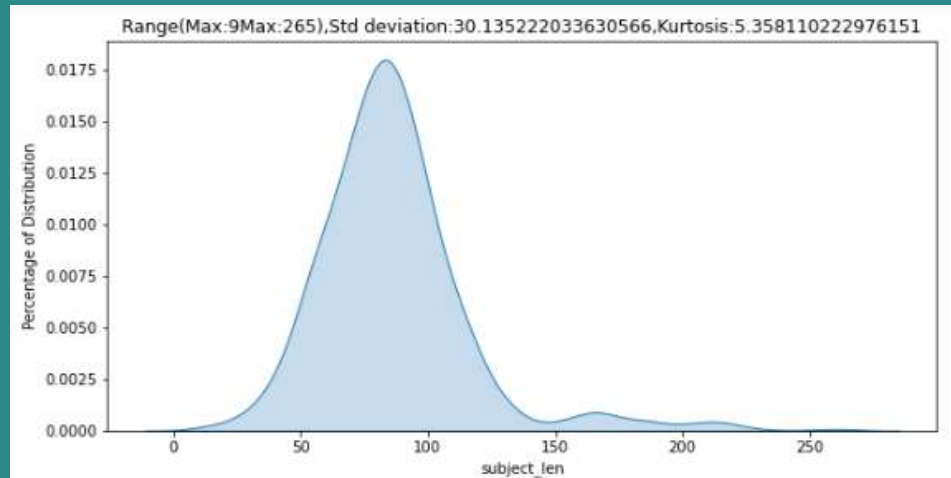
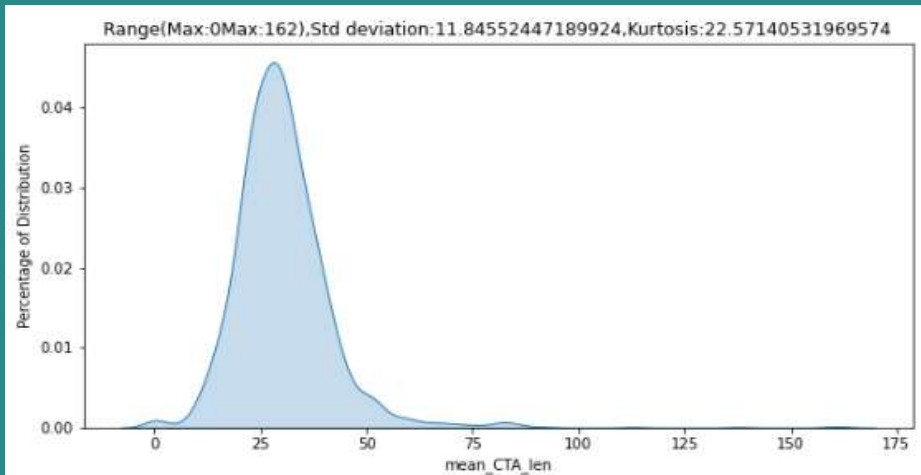


EDA - Univariate Analysis - Categorical



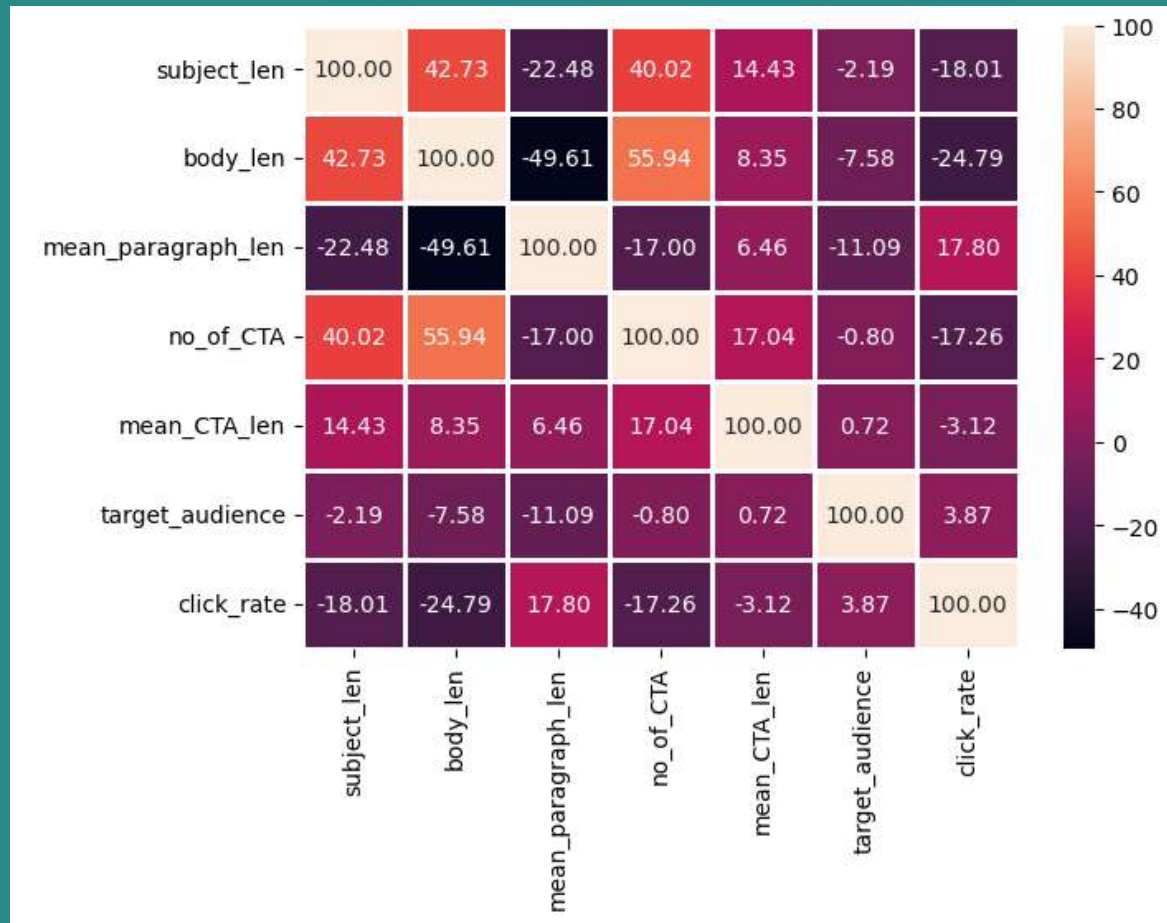
Detailed Analysis and Explanation in Notebook

Univariate Analysis - Numerical



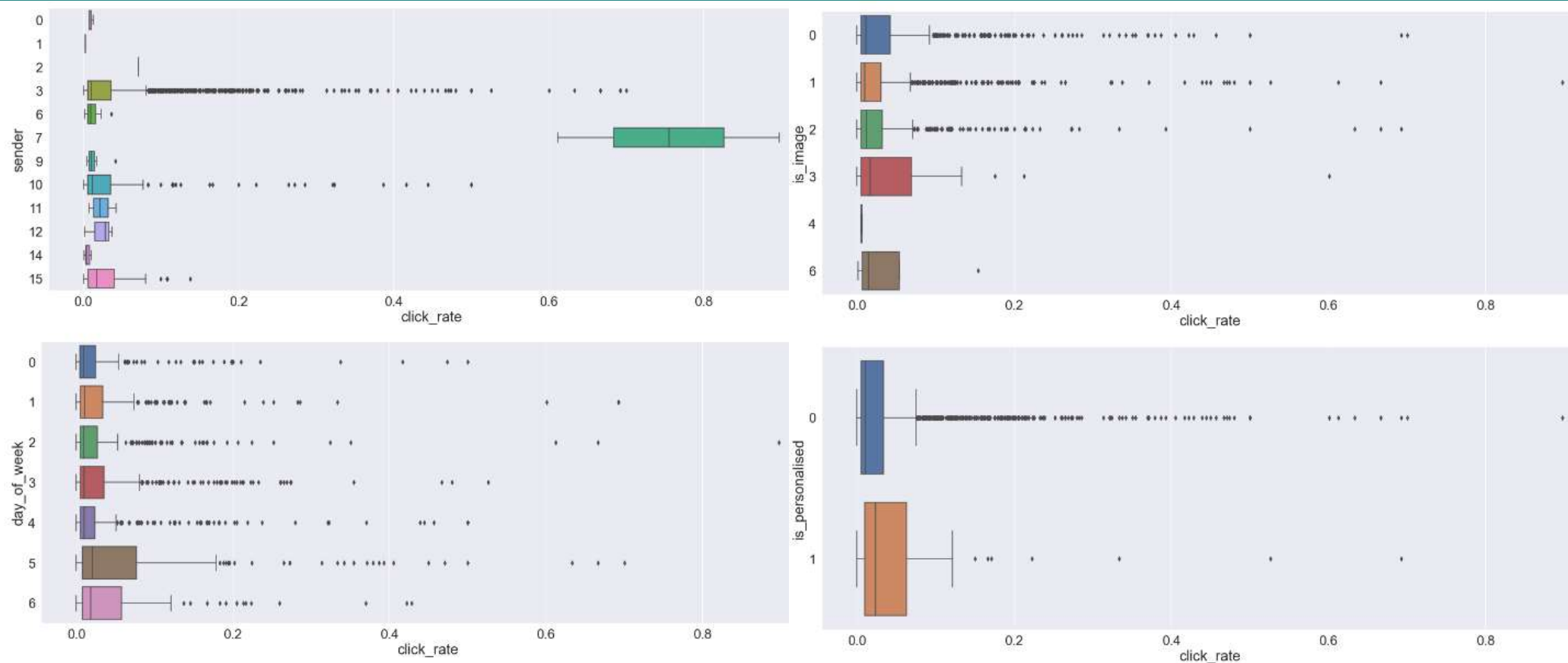
Detailed Analysis and Explanation in Notebook

Bivariate Analysis : Numerical-Numerical



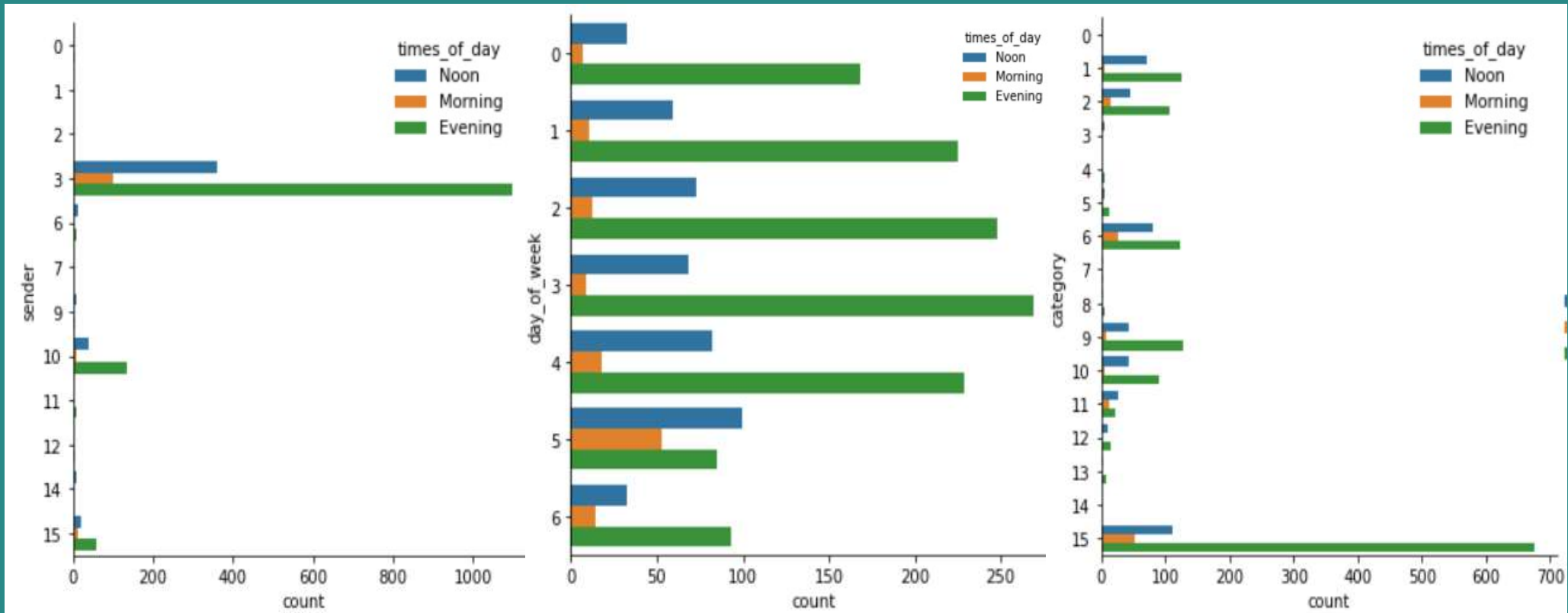
Detailed Analysis and Explanation in Notebook

Bivariate Analysis: Categorical to Numerical



Detailed Analysis and Explanation in Notebook

Bivariate Analysis : Categorical-Categorical



Detailed Analysis and Explanation in Notebook

One – Hot Encoding the Categorical column

- Using the Code - `train=pd.get_dummies(train)`

times_of_day	times_of_day_Evening	times_of_day_Morning	times_of_day_Noon
Noon	0	0	1
Morning	0	1	0
Noon	0	0	1
Evening	1	0	0
Noon	0	0	1

Converting the Columns to Boolean flag

```
2 train.nunique()

sender          12
subject_len    170
body_len       1568
mean_paragraph_len 121
day_of_week     7
is_weekend      2
times_of_day    3
category       16
product        43
no_of_CTA      40
mean_CTA_len   79
is_image        6
is_personalised 2
is_quote        7
is_timer        1
is_emoticons    6
is_discount     2
is_price        8
is_urgency      2
target_audience 17
click_rate     1810
dtype: int64
```

- Changed image column to Boolean to Check only if there is a image or not in a Email
- Changed Quote column to Boolean to Check only if there is a Quote or not in a Email
- Changed Emoticons column to Boolean to Check only if there is a Emoticons or not in a Email
- Changed Price column to Boolean to Check only if there is a Price or not in a Email

Influence of the Change has impacted great on Predication
-Major thing is this about feature has **max of '0' class**
so Converted to Boolean

Dropped Unwanted Features :

- Dropped Is_timer because the class only consist of only one class '0' so dropped Is_timer from the data
- Dropped Is_weekend Because that is Correlated to column and week of the Day .

Splitting the Data- Training and Validation set

- Using `train_test_split` from sklearn library
- Splitting the Training Data to 75% as Training set and 25% as Validation set

Metrics used to for Evaluation

- Using `r2_score` from sklearn library

Model Selection - Approach

- Selecting a XG Booster model
- Then using the XG Booster as a Base model in AdaBoost Model
- Tuning parameters of AdaBoost model manually for Better Results
- Finally the AdaBoost Model Provided with
 - R2 Score of Training set as 0.9964
 - R2 Score of Validation set as 0.6912

Recommendations

Effective Email

- **Category-7** has High Click Rate
- Mail **with image** are having higher click Rate
- **Subject Length** Should between **50 to 100**
- **Body Length** should be **600-1200**
- **mean Length** of Paragraph should be **10-40**.

- What should your subject line look like?

Subject line Should shorter than 100 letter and greater than 50 letters

- What should be the length of the email?

The length of the Mail should be 1250 letters including the Subject

- Do you need images in your email template?

The Click Rate is High with Images



Thank you