ANALYZING GOOGLE APP STORE DATASET

Detailed Project Reports

Presentation By : Siddharth Bhat

PROJECT DETAILS

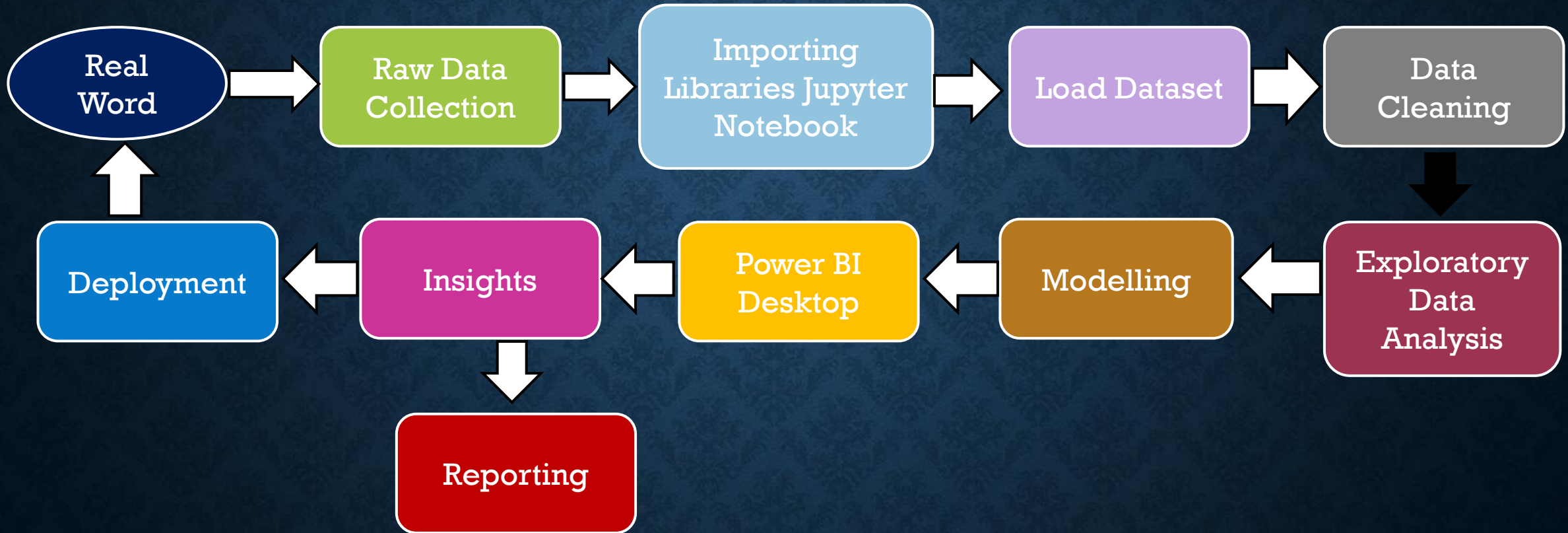| Project Title	Analysing Google App Store Dataset
Technology	Business Intelligence
Domain Name	E - Commerce
Project Difficulties Level	Advanced
Tools	Jupyter Notebook, Power BI, Ms-Excel, Power Point

OBJECTIVE

Find key metrics and factors and show the meaningful relationships between attributes.

PROBLEM STATEMENT

❑ Load the dataset and find for me some information from this large data. This is form of data mining

❑ What all information can be achieved by mining this data, would be brainstormed by the interns and factors and show the meaningful relationships between attributes.

❑ Do your own research and come up with your findings.
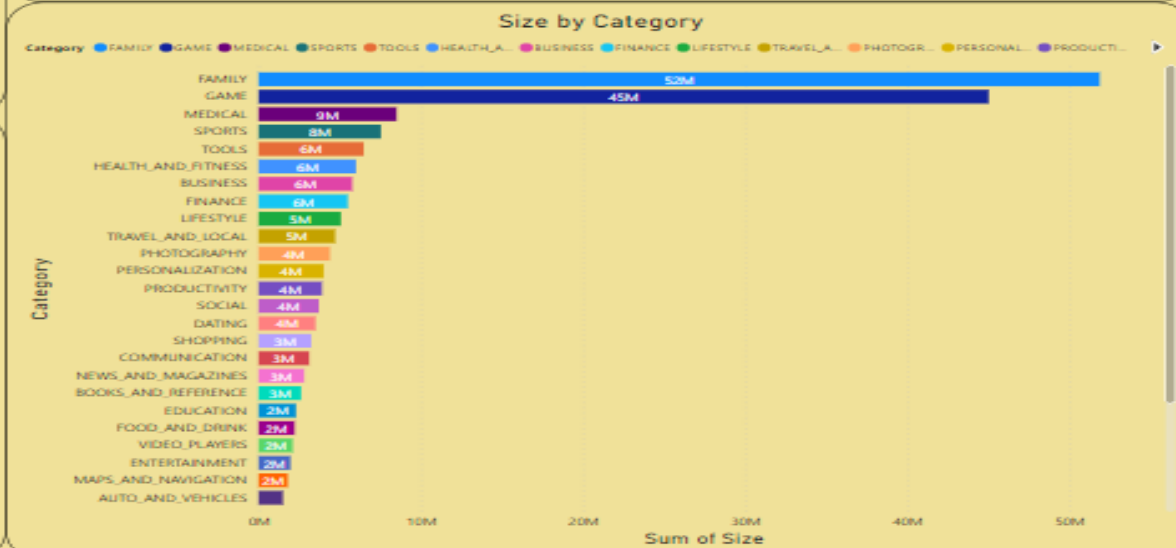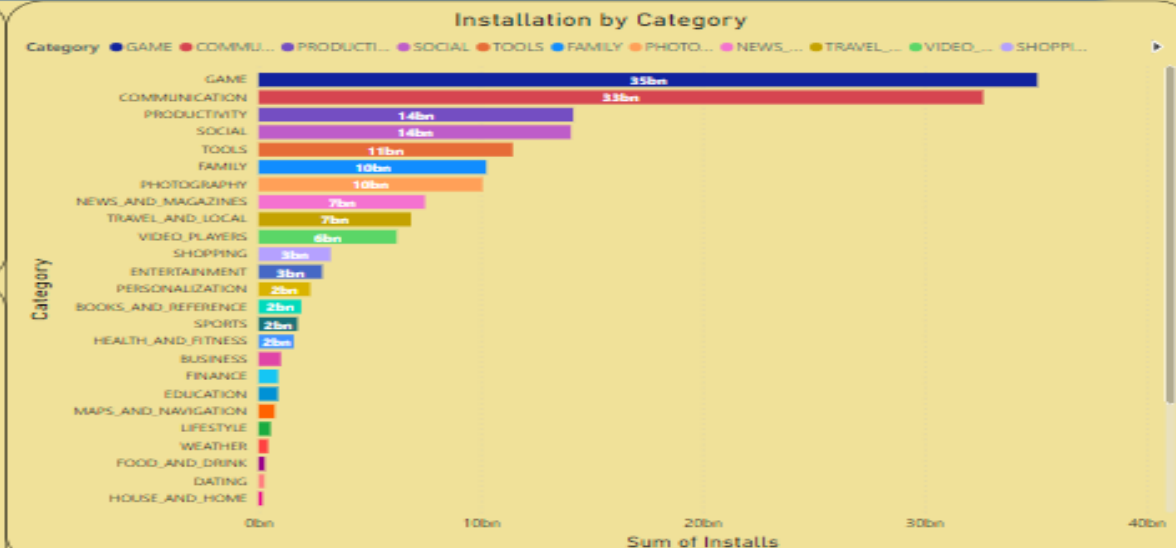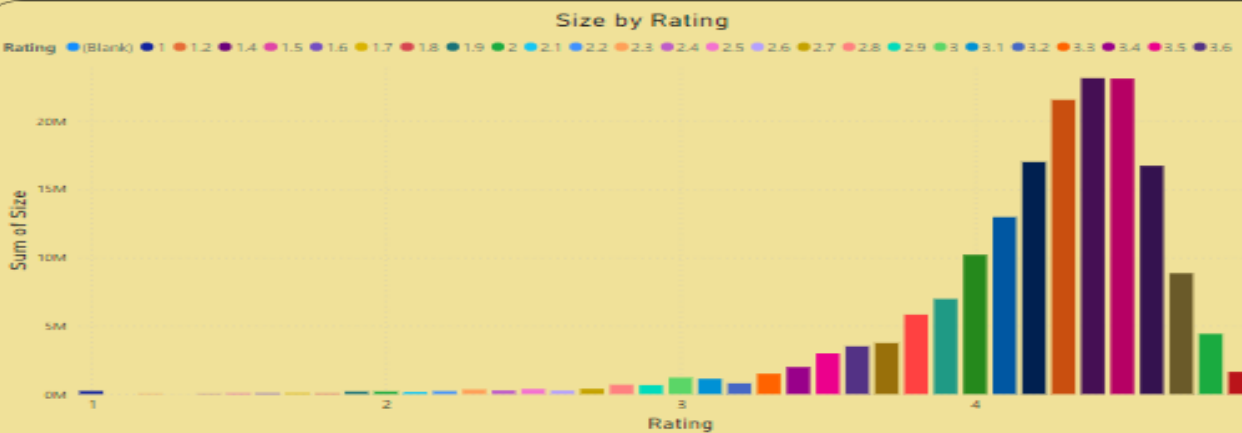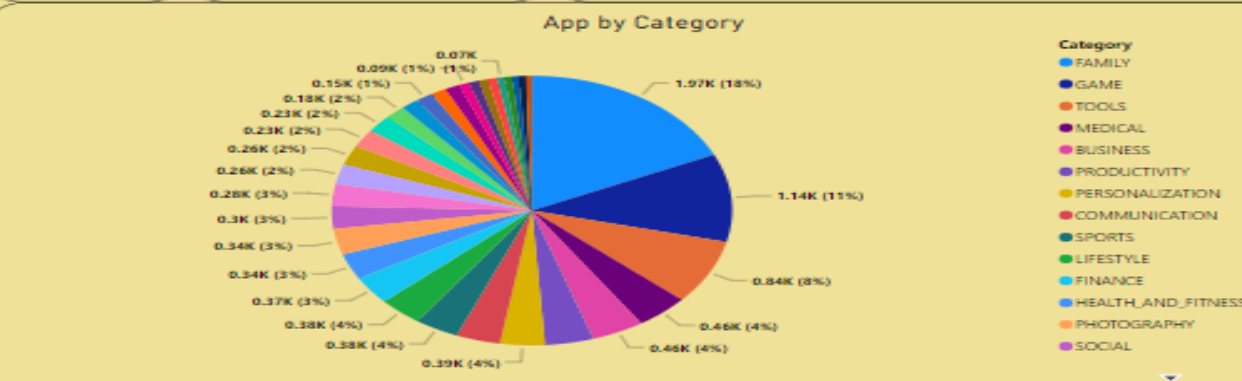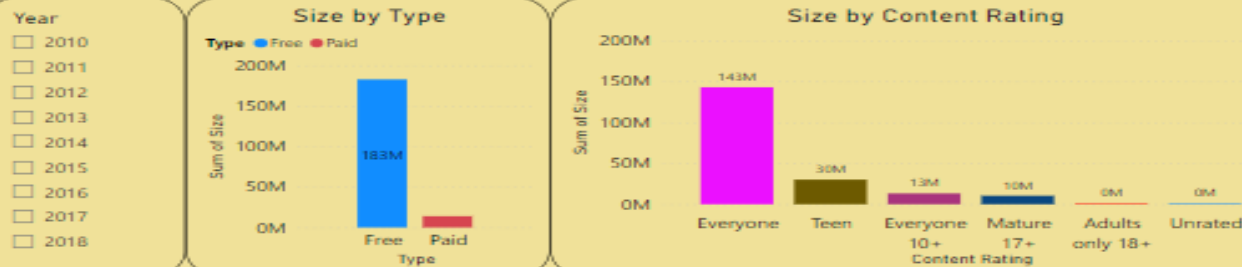
ARCHITECTURE

DATASET INFORMATION

This is a Google App Store dataset that contains Information like App, Category, Rating, Review, Size, Installation, Type
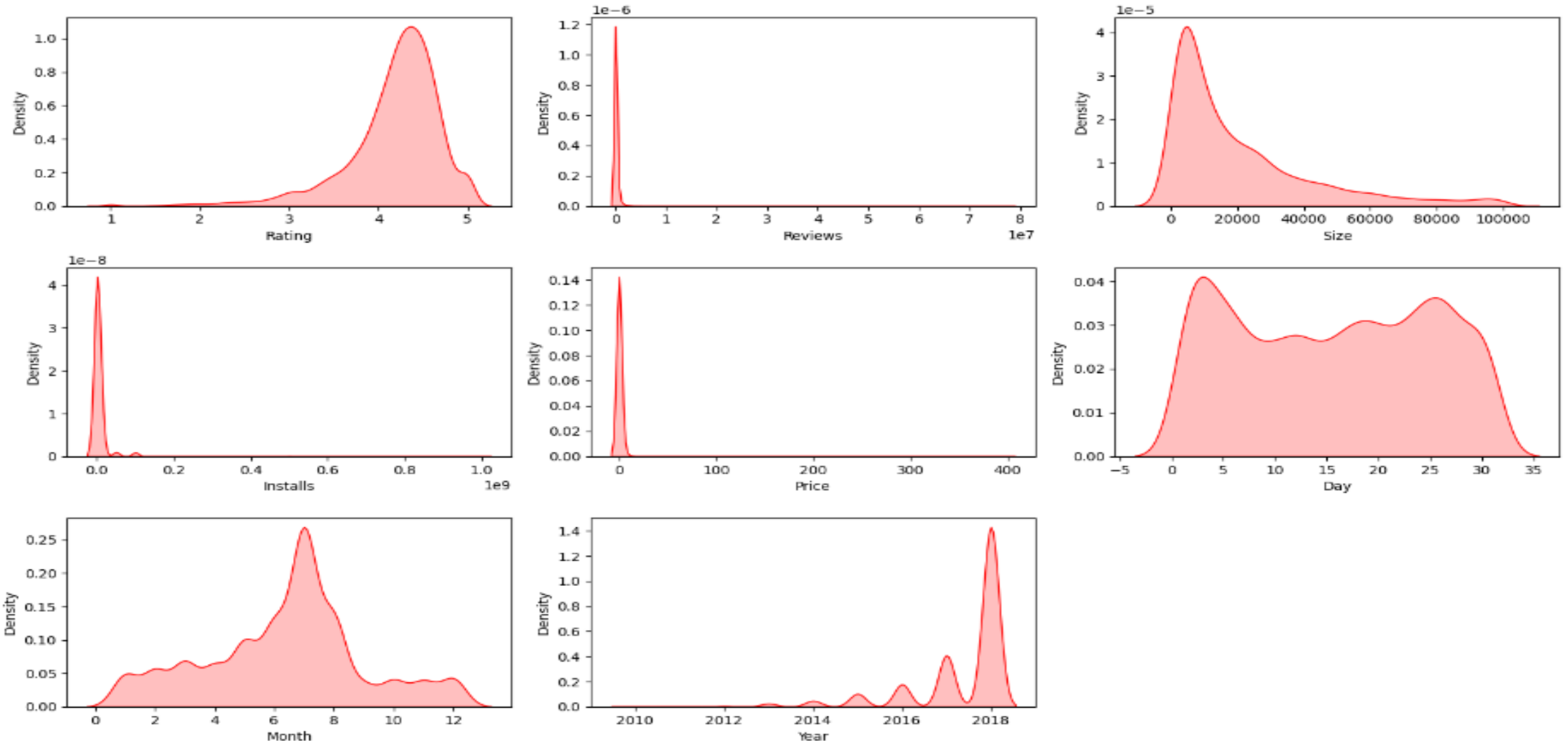
DASHBOARD

Analysing Google App Store Dashboard
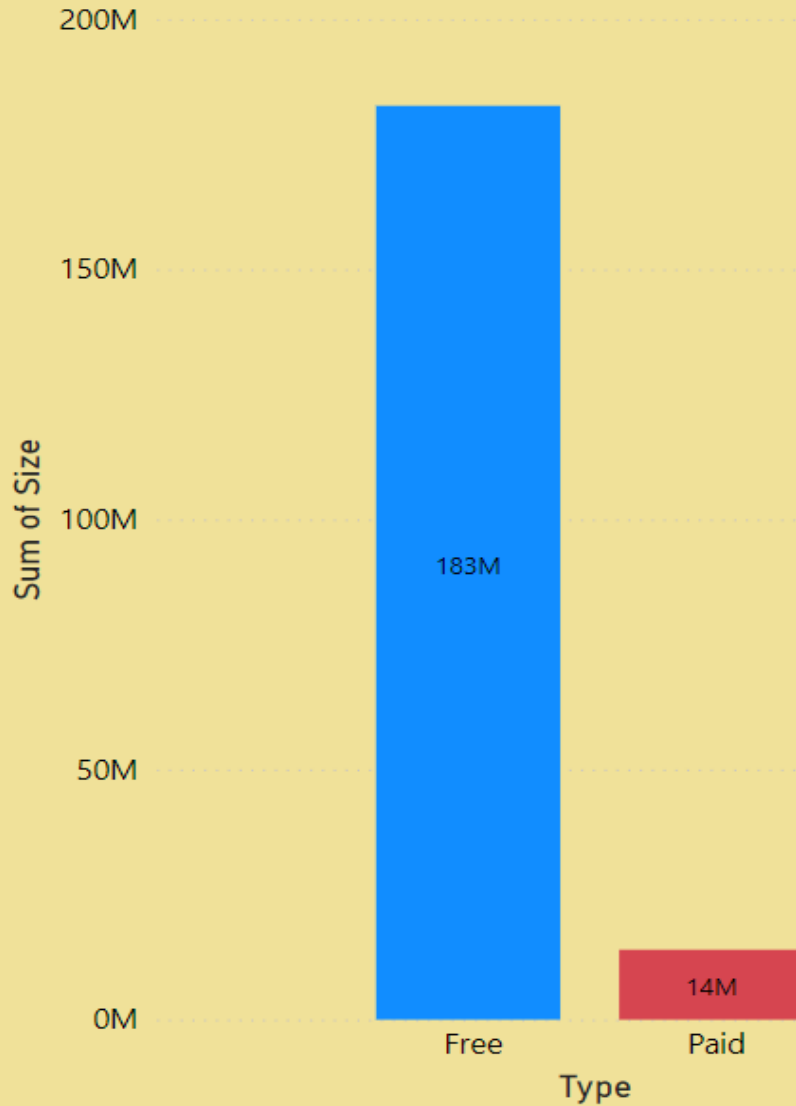
Univariate Analysis of Numerical Features

Univariate Analysis of Categorical Features
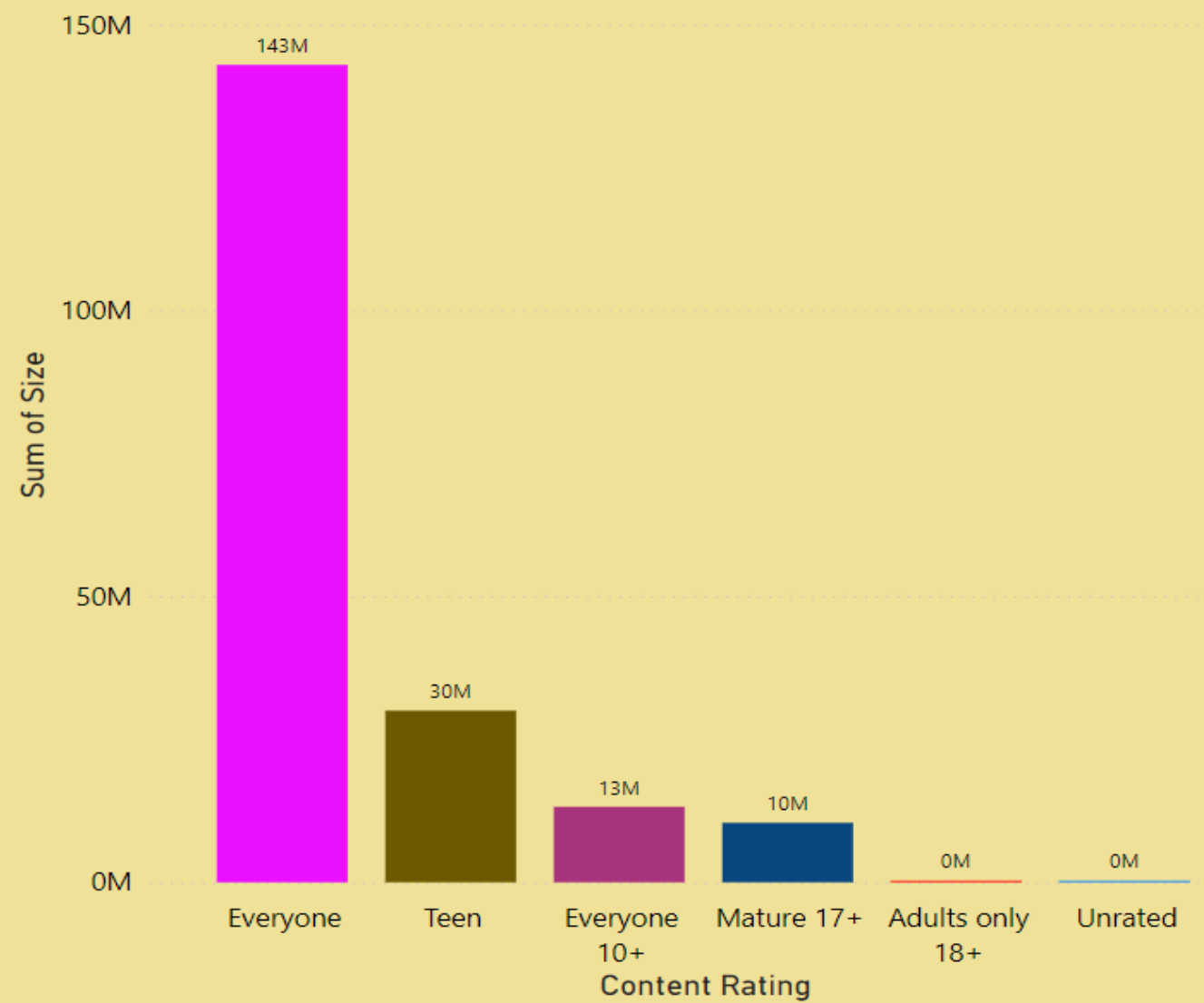
Install Category wise Application

App by Category

Category
- FAMILY
- GAME
- TOOLS
- MEDICAL
- BUSINESS
- PRODUCTIVITY
- PERSONALIZATION
- COMMUNICATION
- SPORTS
- LIFESTYLE
- FINANCE
- HEALTH_AND_FITNESS
- PHOTOGRAPHY
- SOCIAL
- NEWS_AND_MAGAZINES
- SHOPPING
- TRAVEL_AND_LOCAL
- DATING
- BOOKS_AND_REFERENCE
- VIDEO_PLAYERS
- EDUCATION
- ENTERTAINMENT
- MAPS_AND_NAVIGATION
- FOOD_AND_DRINK
- HOUSE_AND_HOME

1.97K (18%)
1.14K (11%)
0.84K (8%)
0.46K (4%)
0.46K (4%)
0.42K (4%)
0.39K (4%)
0.39K (4%)
0.38K (4%)
0.38K (4%)
0.37K (3%)
0.34K (3%)
0.34K (3%)
0.3K (3%)
0.28K (3%)
0.26K (2%)
0.26K (2%)
0.23K (2%)
0.23K (2%)
0.18K (2%)
0.16K (1%)
0.14K (1%)
0.09K (1%)
0.08K (1%)
0.06K (1%)

Top Application Installation in Size
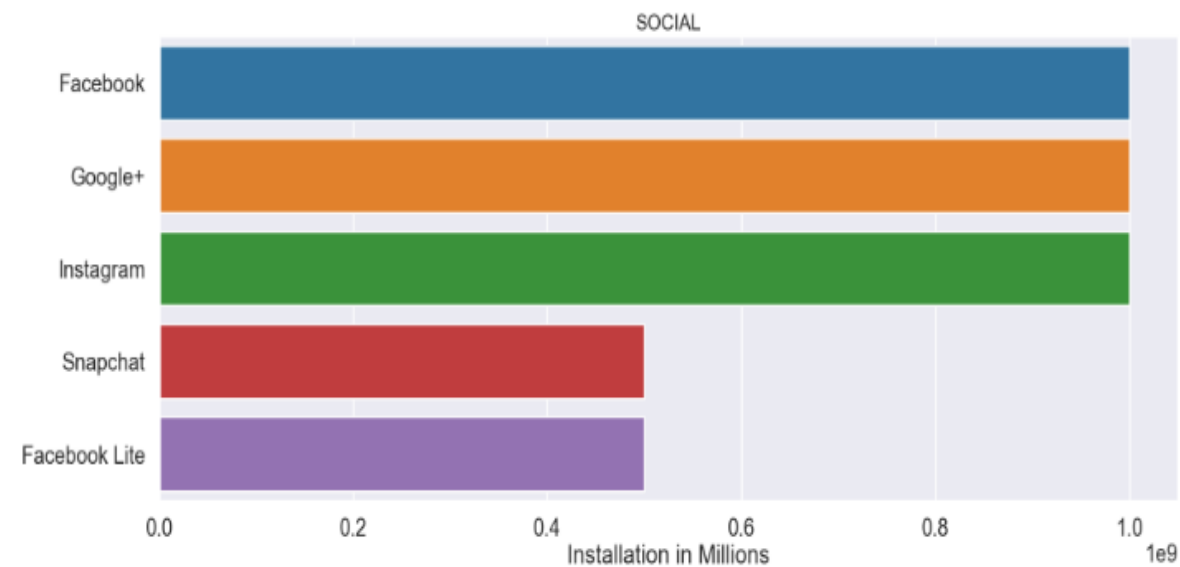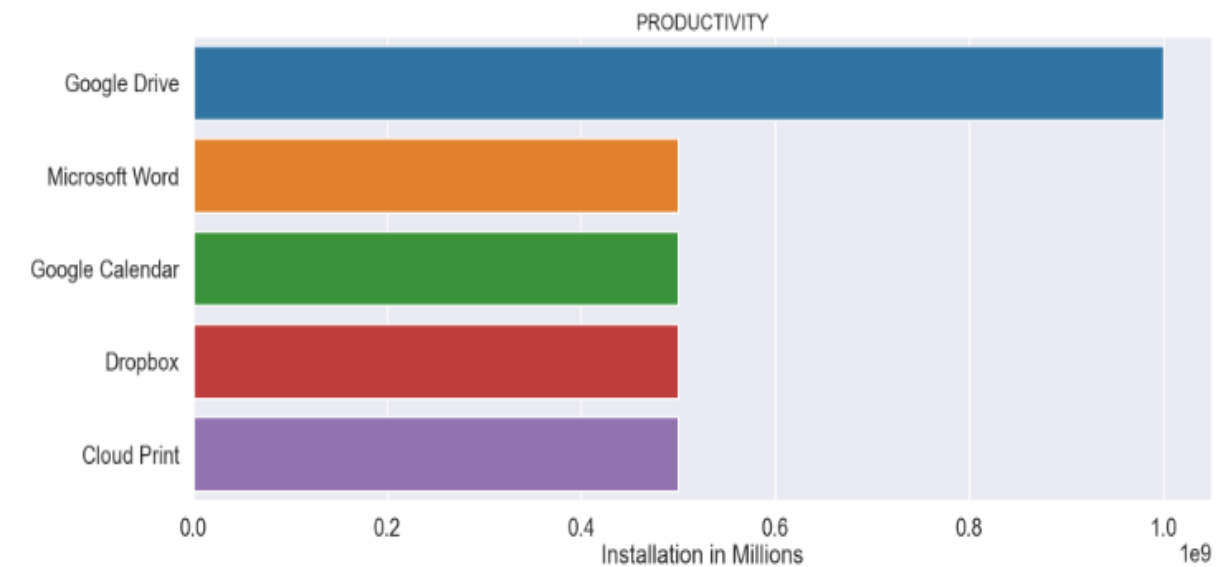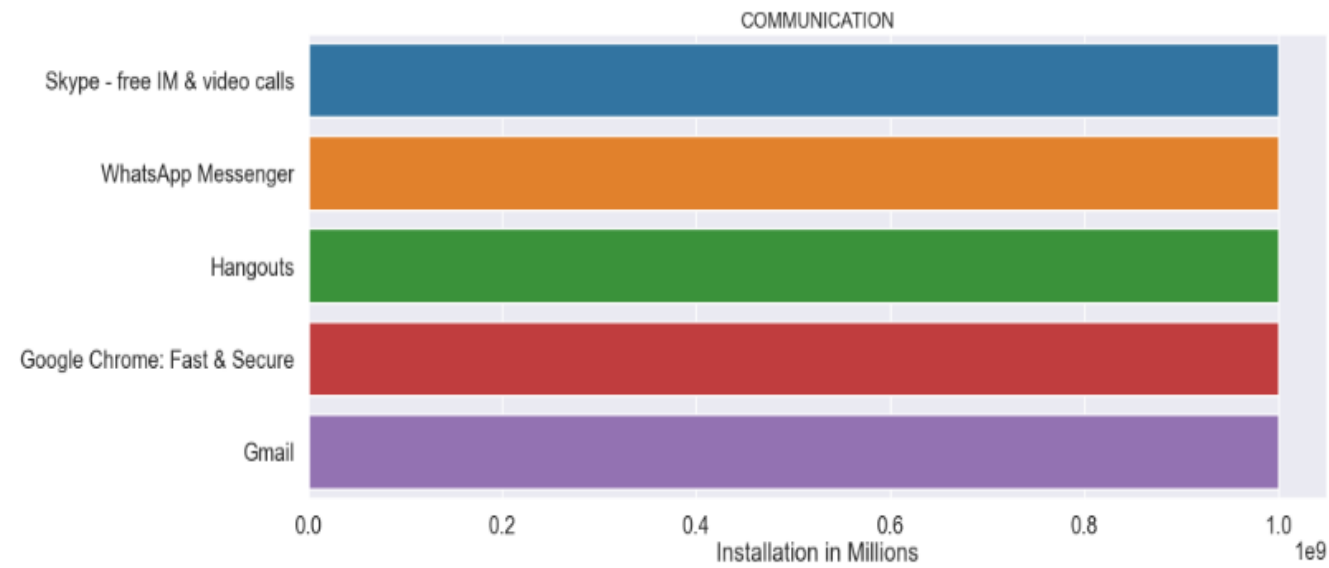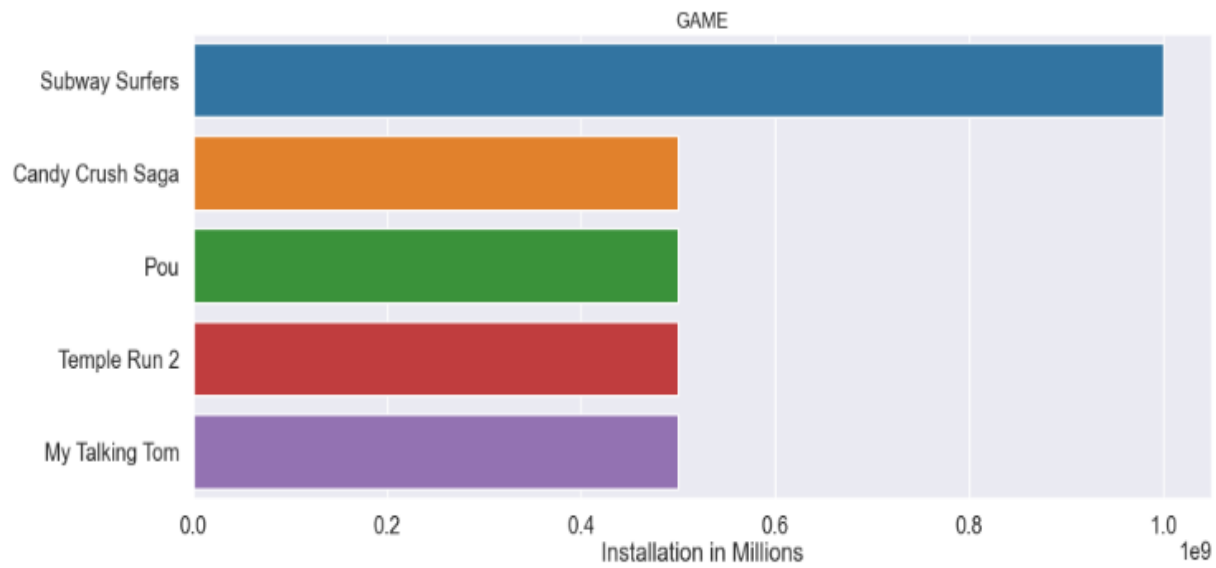
Category wise Rating
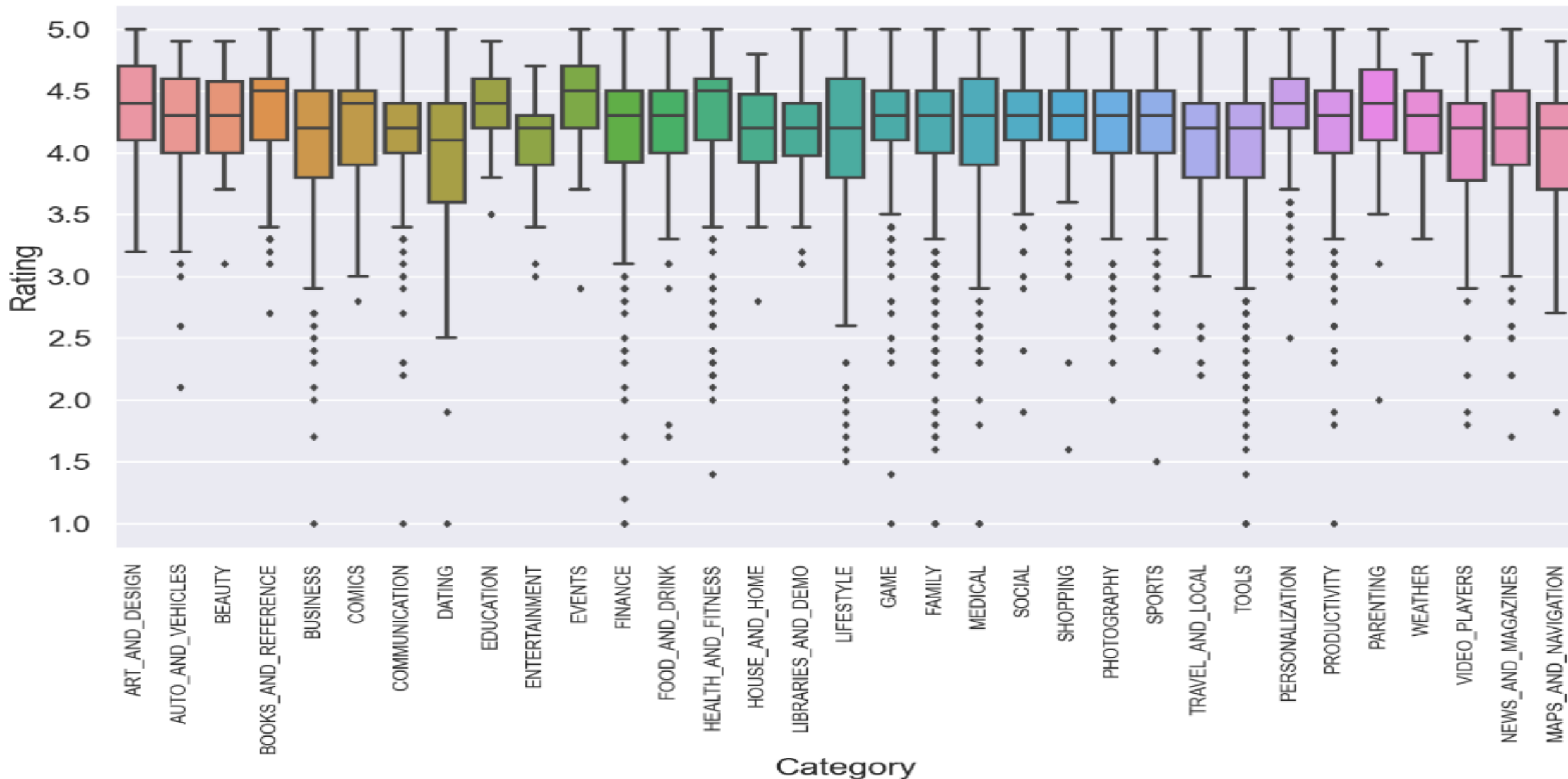
Installation wise Rating

Category wise Installation

Installation by Category

Category ● GAME ● COMM... ● PRODU... ● SOCIAL ● TOOLS ● FAMILY ● PHOTO... ● NEWS_... ● TRAVEL_... ● VIDEO_... ● SHOPPI... ● ENTERT... ● PERSON... ● BOOKS_... ● SPORTS ● HEALTH_...

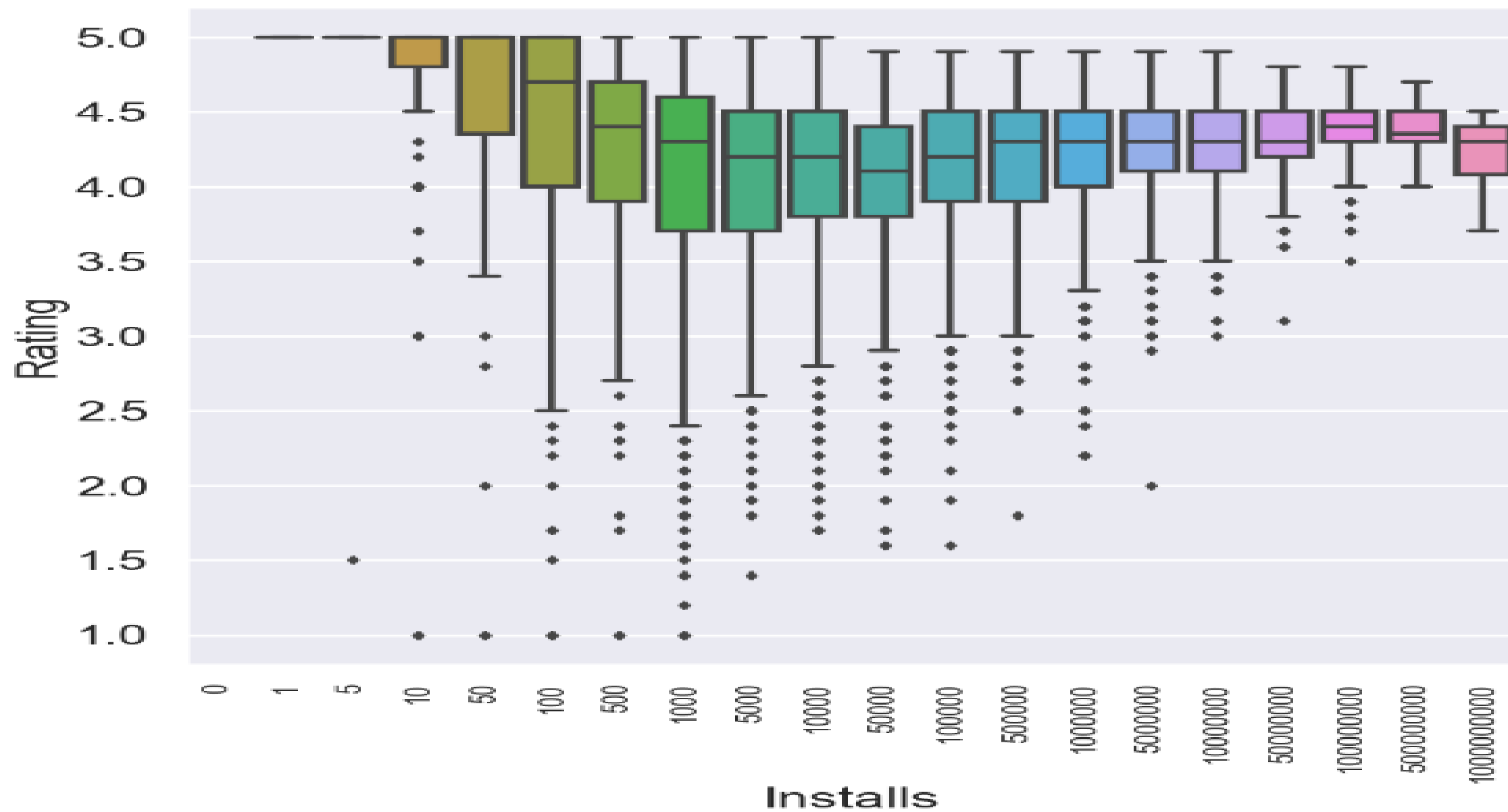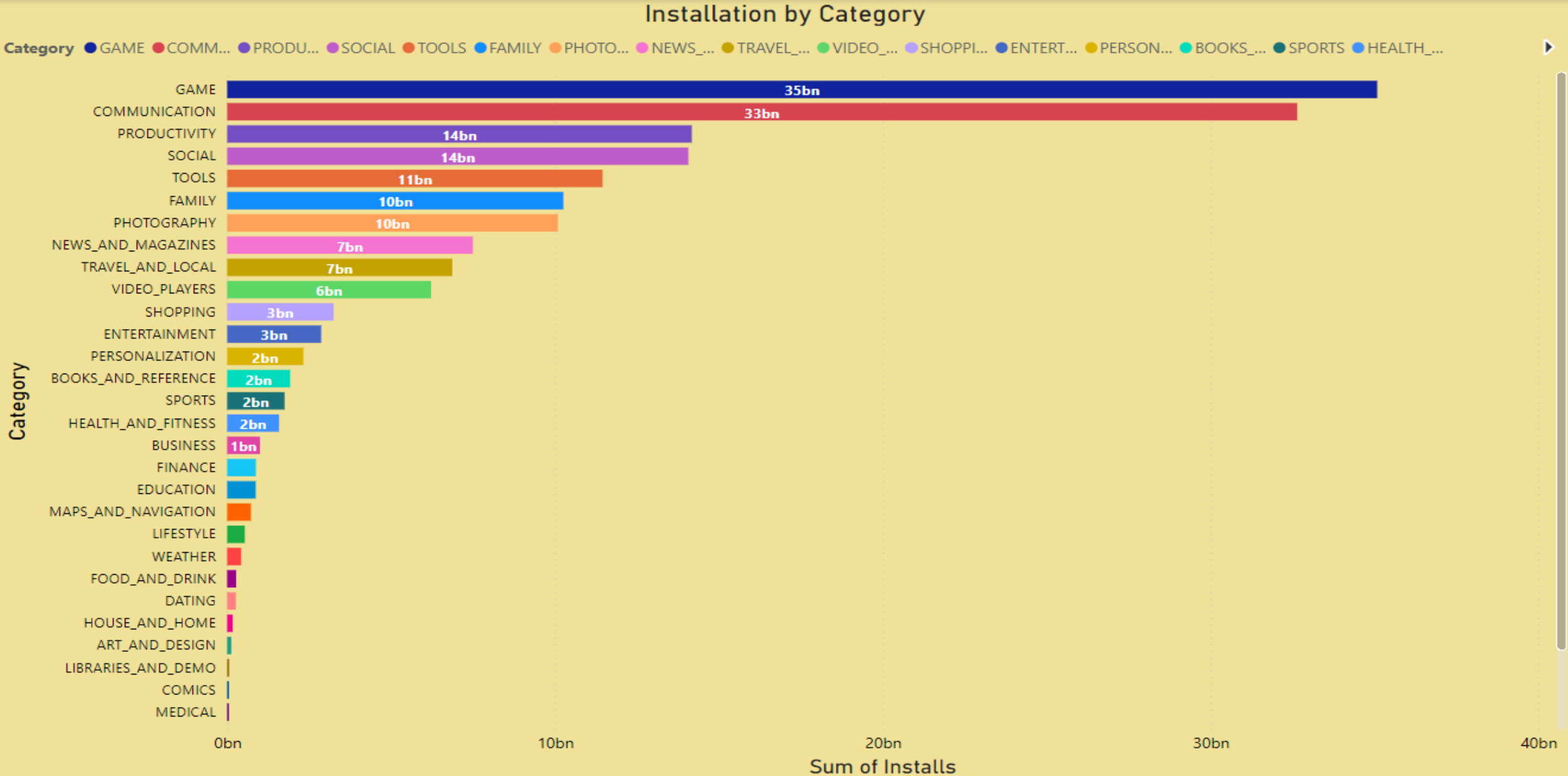| Category	Sum of Installs
GAME	35bn
COMMUNICATION	33bn
PRODUCTIVITY	14bn
SOCIAL	14bn
TOOLS	11bn
FAMILY	10bn
PHOTOGRAPHY	10bn
NEWS_AND_MAGAZINES	7bn
TRAVEL_AND_LOCAL	7bn
VIDEO_PLAYERS	6bn
SHOPPING	3bn
ENTERTAINMENT	3bn
PERSONALIZATION	2bn
BOOKS_AND_REFERENCE	2bn
SPORTS	2bn
HEALTH_AND_FITNESS	2bn
BUSINESS	1bn
FINANCE	
EDUCATION	
MAPS_AND_NAVIGATION	
LIFESTYLE	
WEATHER	
FOOD_AND_DRINK	
DATING	
HOUSE_AND_HOME	
ART_AND_DESIGN	
LIBRARIES_AND_DEMO	
COMICS	
MEDICAL	

0bn   10bn   20bn   30bn   40bn

Sum of Installs

Size wise Rating

Size by Rating

Application wise Review

Reviews by App — Sum of Reviews by App
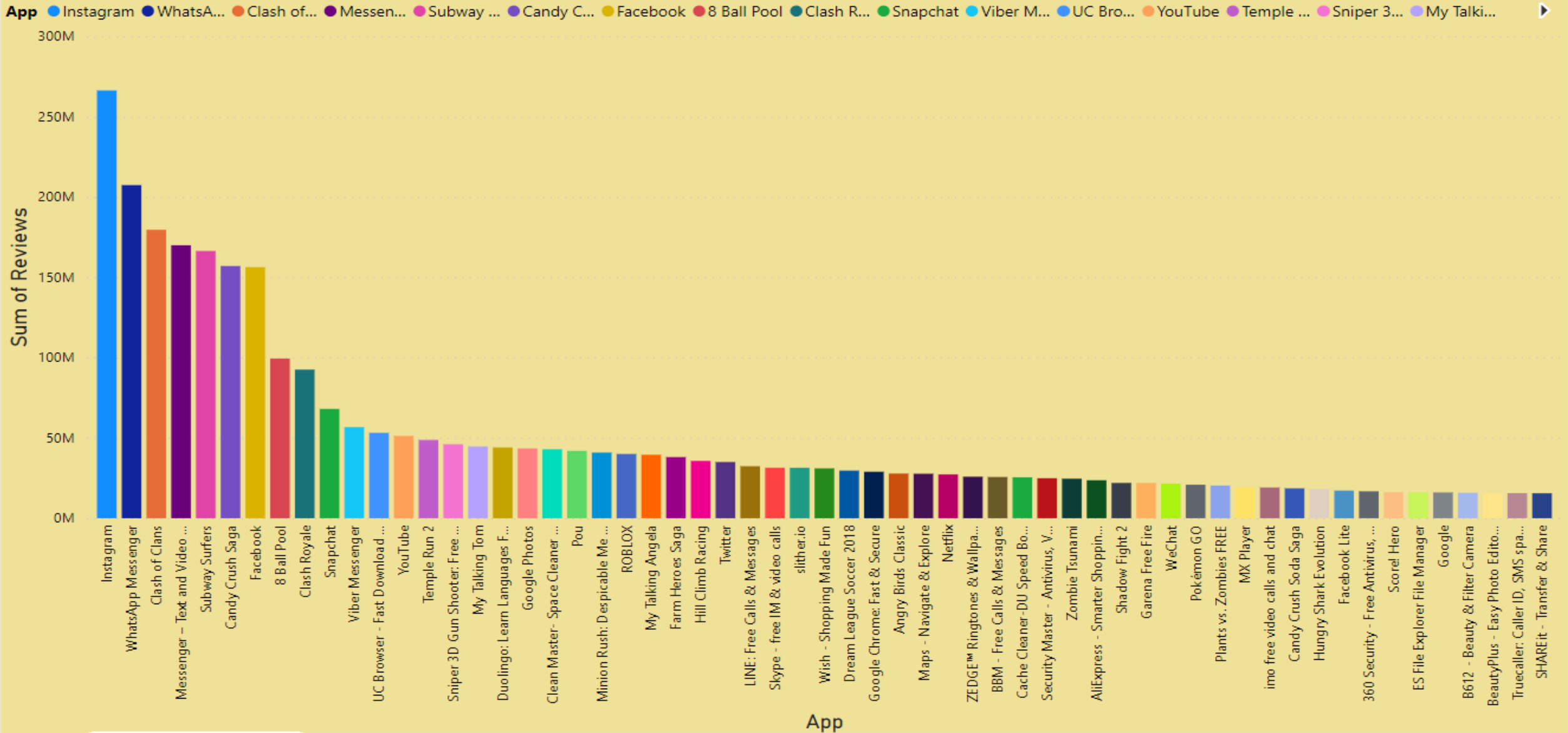
Type Wise App And Installation

App by Type

0K (0.01%)
0.8K (7.38%)
10.04K (92.61%)

Type
- Free
- Paid
- (Blank)

Installation by Type

0bn (0.04%)
168bn (99.96%)

Type
- Free
- Paid
- (Blank)
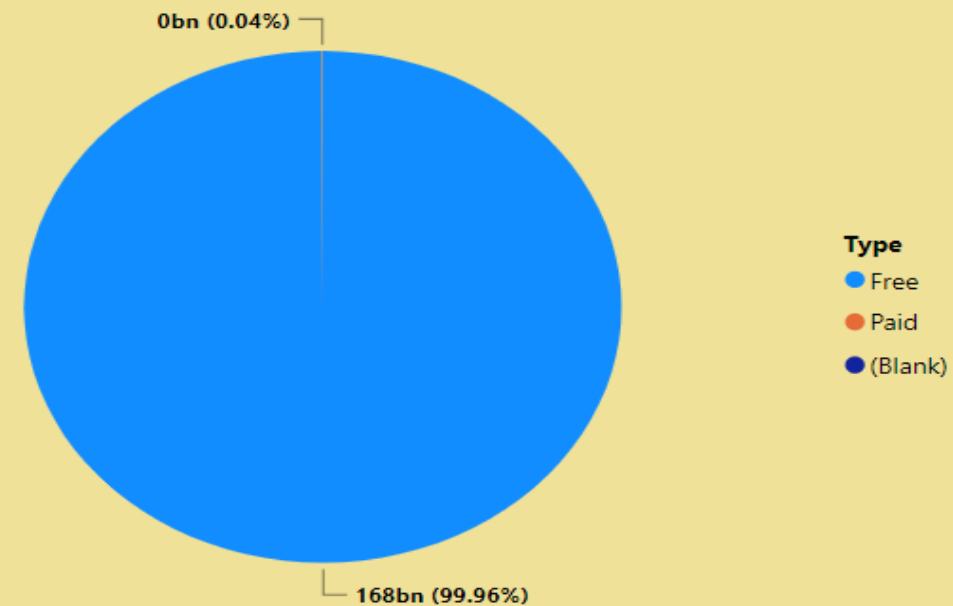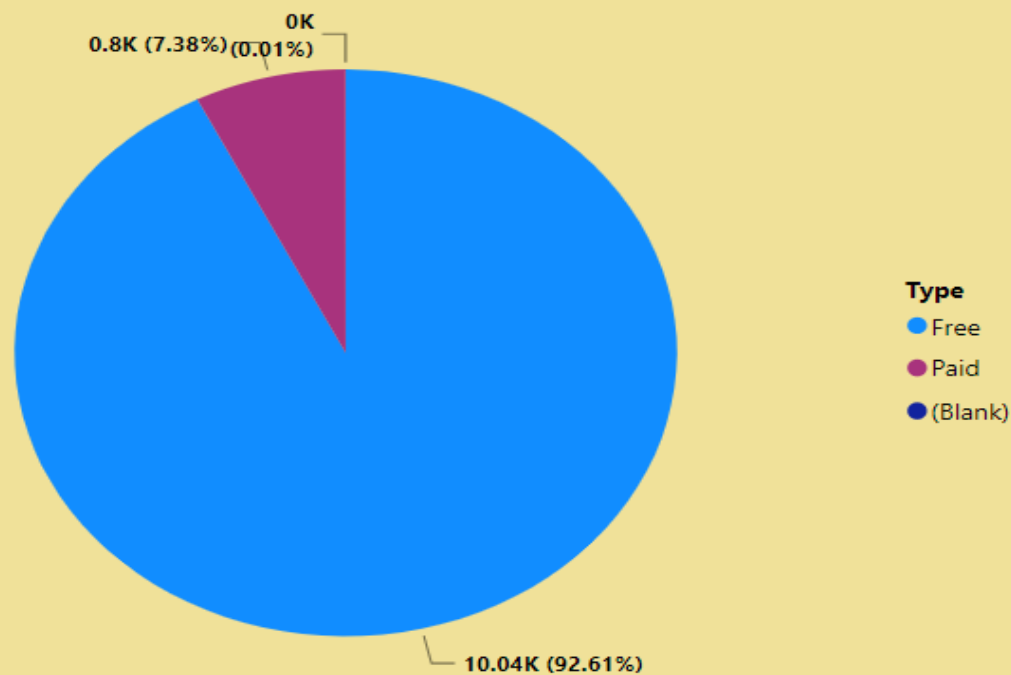
QUESTIONS AND ANSWERS

Q1) What's the source of data?
Ans) The Dataset was taken from iNeuron's Provided Project Description Document.
https://drive.google.com/file/d/1QpHsrW3XudFfarTD-e8jnMjGubdRPgvS/view

Q2)What was the type of data?
Ans) The data was a combination of numerical and Categorical values.

Q 3)What was the complete flow you followed in this Project?
Ans) Refer slide 7th for better Understanding

Q4) What techniques were you using for data?
Ans) Removing unwanted attributes.
    Visualizing relation of independent variables with each other and output variables.
    Checking and changing distribution of continuous values.
    Removing outliers and Cleaning data are present.
    Transforming data to yield the desired result.

THANK YOU