**DATA SCIENCE**

**MINI PROJECT**

# Google Play St Google Play Store Apps

# Google Google Play Store Apps

# Ratings Prediction

**Cross Sell Prediction**

**GROUP NO: 17**

**GROUP MEMBERS:**

|  |  |  |  |
| --- | --- | --- | --- |
| **NAME** | **ROLL NO** | **GR NO** | **BATCH** |
| **Kundan Kabra** | **321023** | **21810729** | **A1** |
| **Ketan Kolte** | **321026** | **21810707** | **A1** |
| **Pranav Sarda** | **321047** | **21810773** | **A2** |
| **Viraj Patil** | **321058** | **21810468** | **A3** |

**PROBLEM STATEMENT:** Identify problem statement. Use Structured data set. Define 3 to 4 objectives.

Perform

**1. Data Interpretation**

**2. Data Pre-processing**

**3. Data visualization**

**4. Data modelling**

**INSPIRATION:** The Play Store apps data has enormous potential to drive app-making businesses to success. Actionable insights can be drawn for developers to work on and capture the Android market!

**OBJECTIVE:**

* Data-Interpretation & Data-Pre-processing for better understanding of data.
* Visualization of the dataset by using various plots and graphs.
* Build a model to predict Google Play Store Apps Ratings**.**
* Comparison between different models used for prediction.

**DATA SET DESCRIPTION:** The data -set we used is **Google Play Store Apps** taken from Kaggle.

The data is collected with the help of Python and Scrapy running on a google cloud vm instance. The data was collected on December 2020.

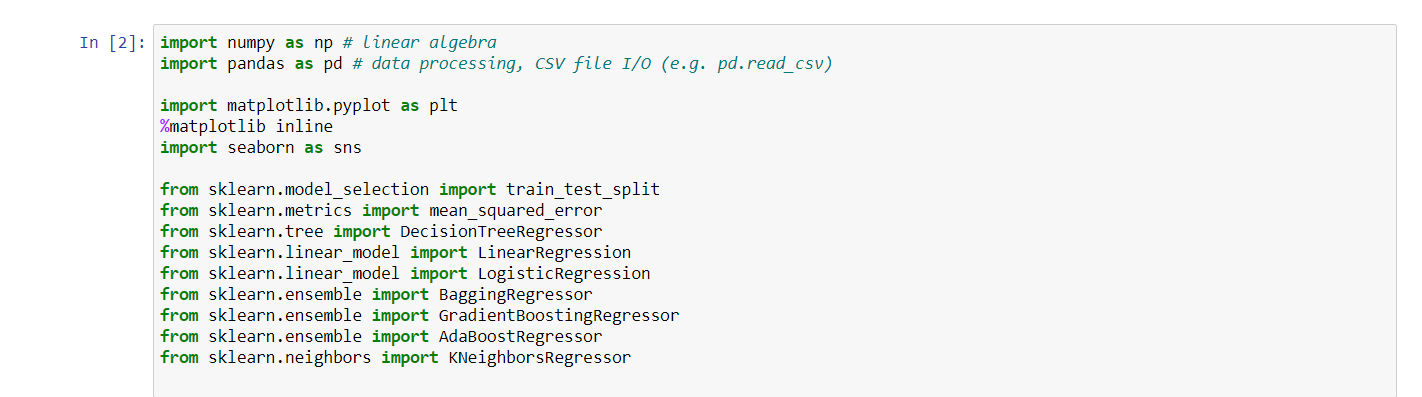
This dataset contains Application data of more than 600K applications with the following 23 attributes.

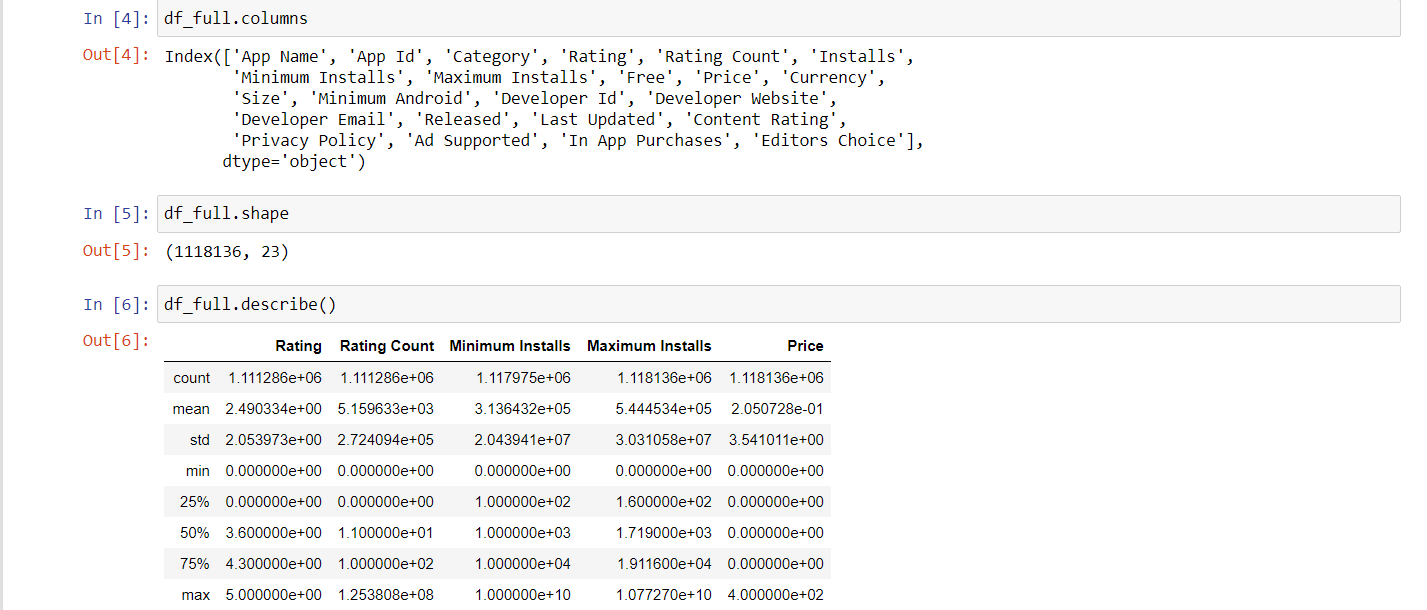
**DATA DICTIONARY:**

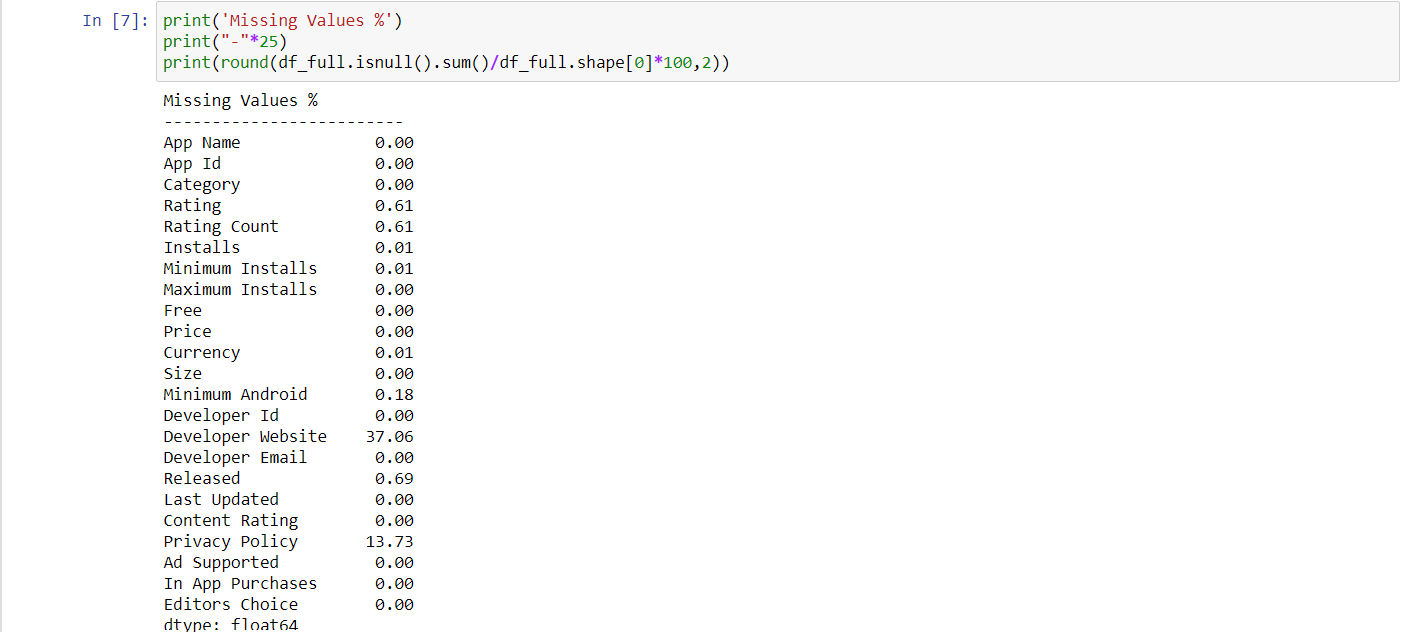
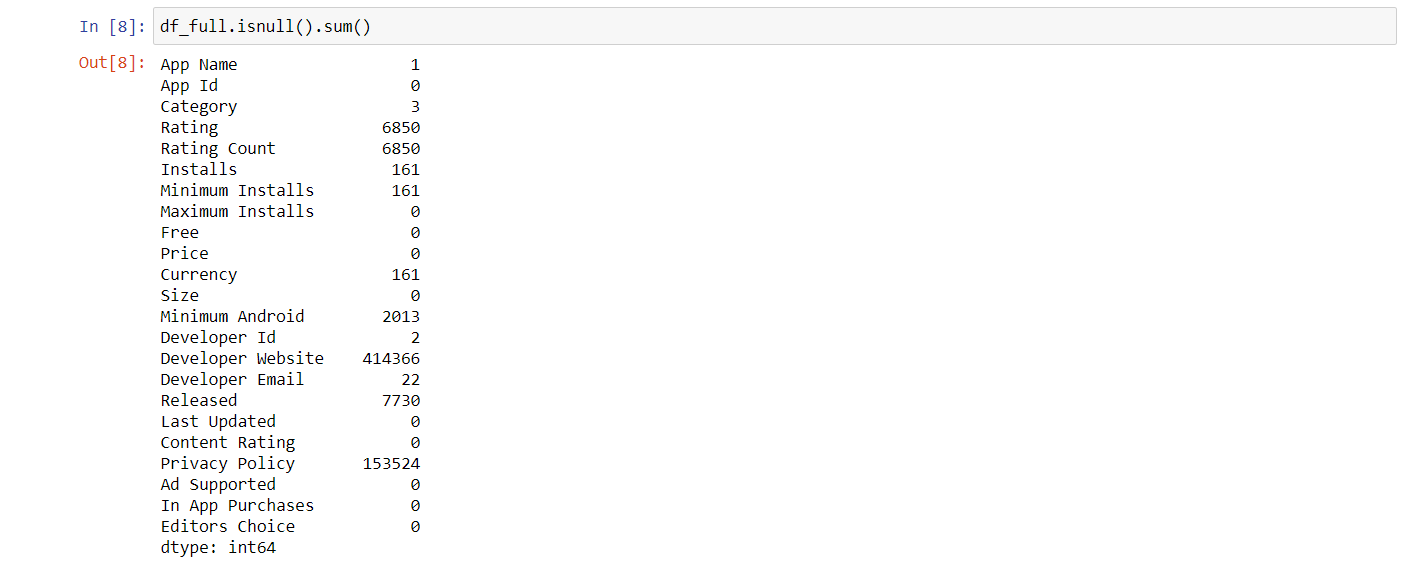
|  |  |  |
| --- | --- | --- |
| **SR NO** | **VARIABLE** | **DEFINITION** |
| 1 | **App Name** | Name of the app |
| 2 | **App Id** | Package name |
| 3 | **Category** | App category |
| 4 | **Rating** | Average rating |
| 5 | **Rating Count** | Number of rating |
| 6 | **Installs** | Approximate install count |
| 7 | **Minimum Installs** | Approximate minimum app install count |
| 8 | **Maximum Installs** | Approximate maximum app install count |
| 9 | **Free** | Whether app is Free or Paid |
| 10 | **Price** | App price |
| 11 | **Currency** | App currency |
| 12 | **Size** | Size of application package |
| 13 | **Minimum Android** | Minimum android version supported |
| 14 | **Developer Id** | Developer Id in Google Playstore |
| 15 | **Developer Website** | Website of developer |
| 16 | **Developer Email** | Email-id of developer |
| 17 | **Released** | App launch date on Google Playstore |
| 18 | **Privacy Policy** | Privacy policy from developer |
| 19 | **Last Updated** | Last app update date |
| 20 | **Content Rating** | Maturity level of app |
| 21 | **Ad Supported** | Ad support in app |
| 22 | **In app purchases** | In-App purchases in app |
| 23 | **Editor Choice** | Whether rated as Editor Choice. |

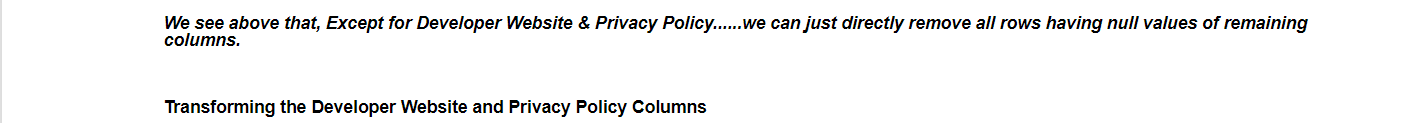
**CODE AND OUTPUT:**

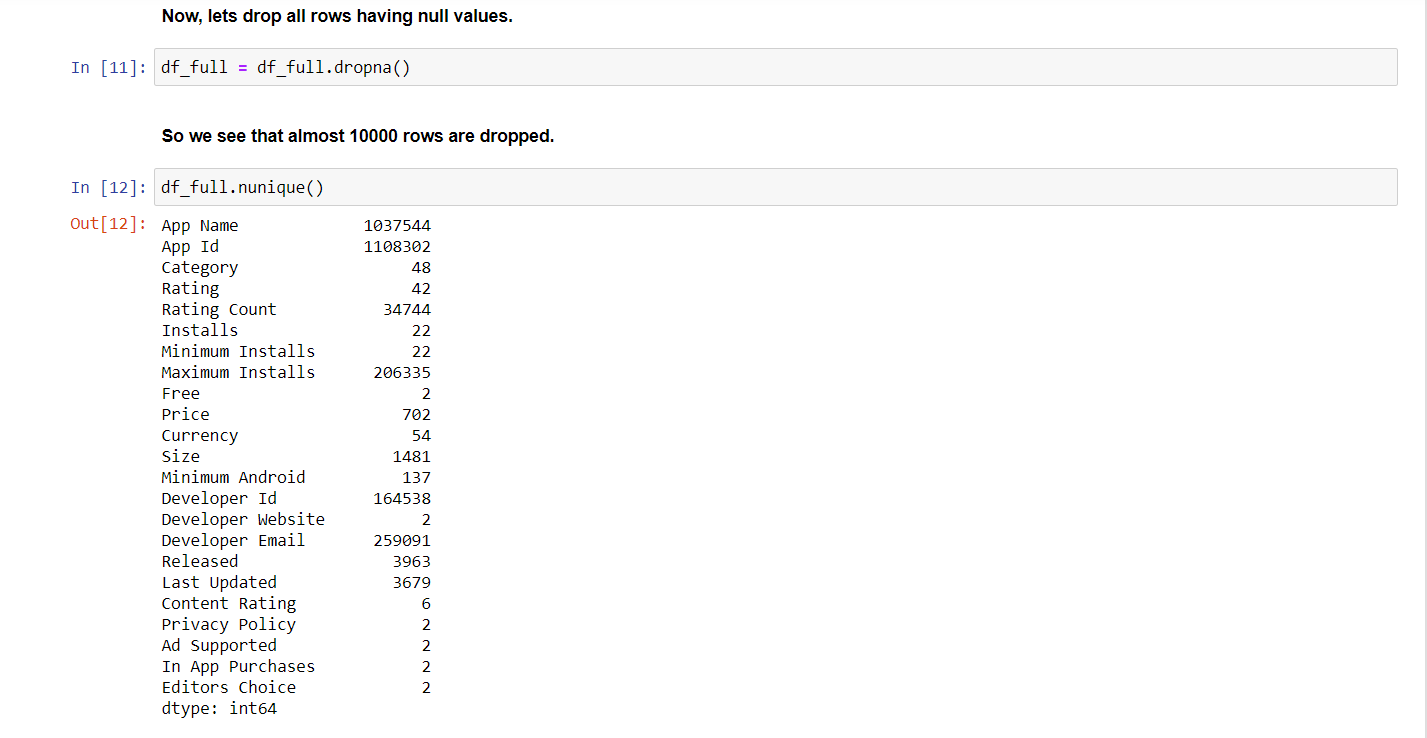
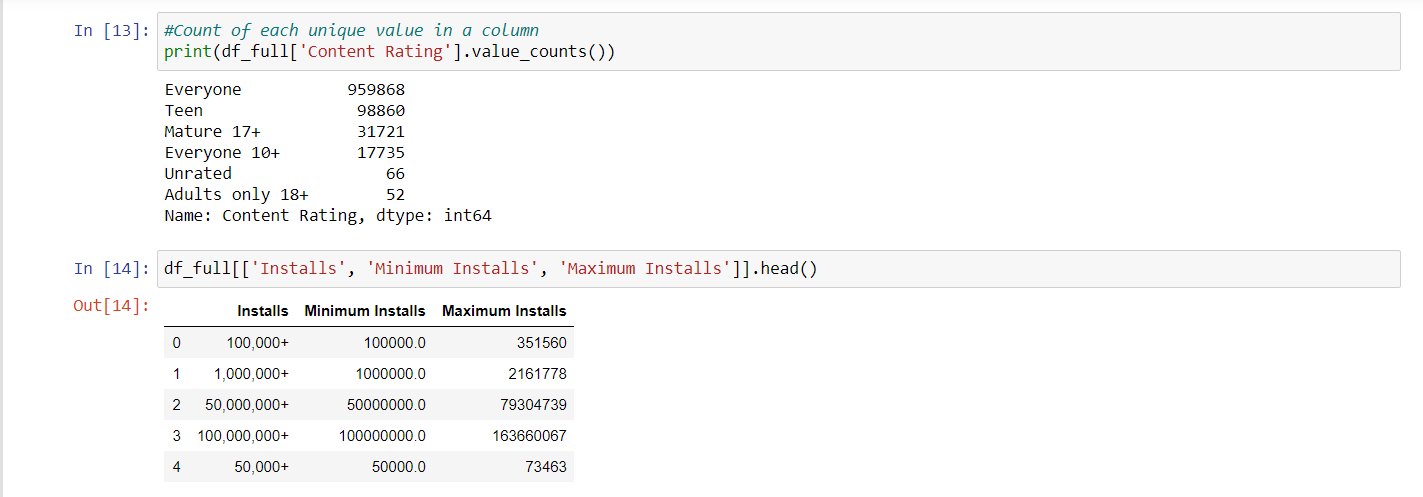
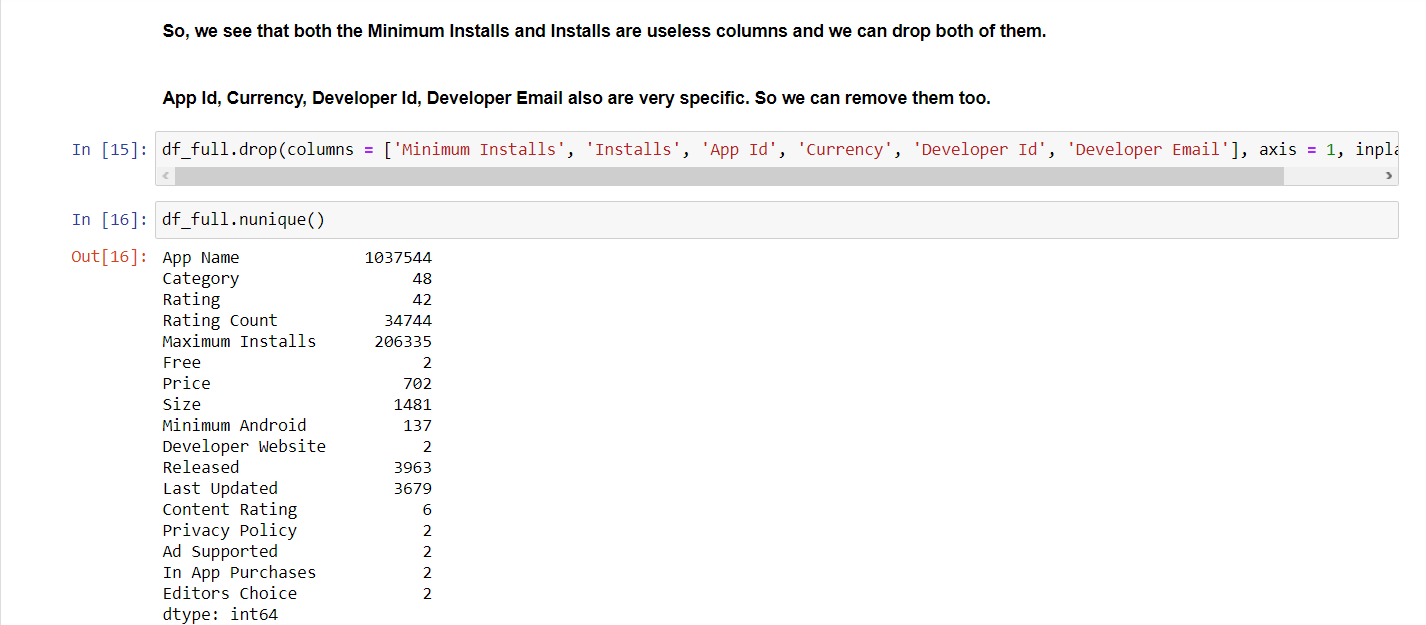
**OBJECTIVE 1:** Data-Interpretation & Data-Pre-processing for better understanding of data.

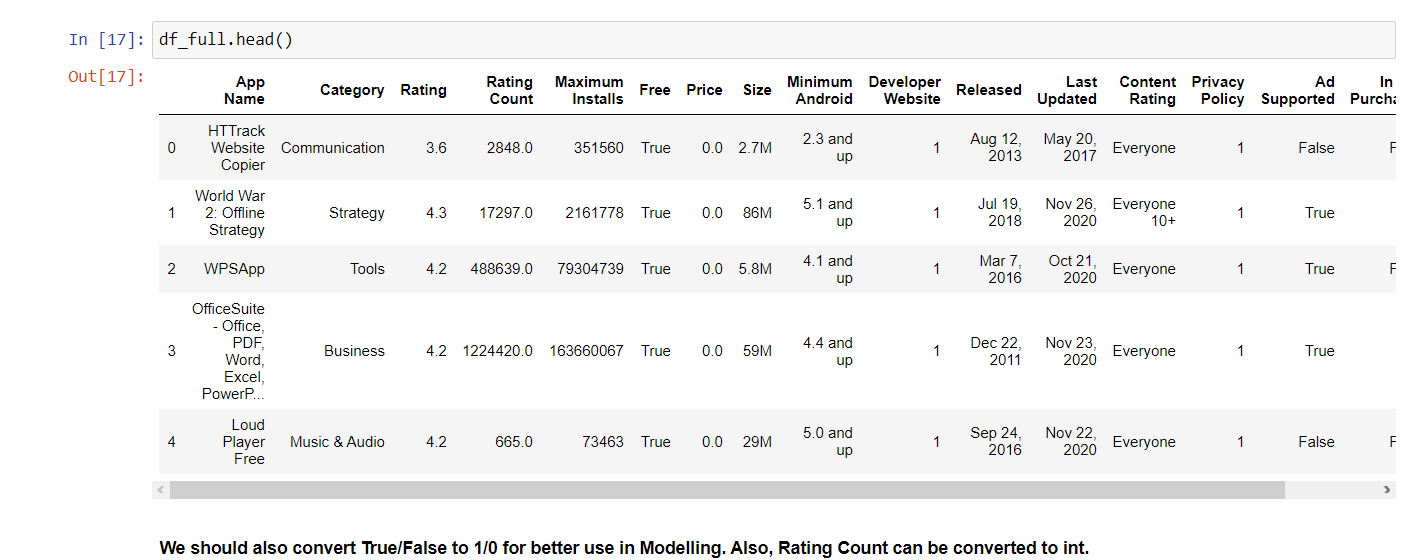
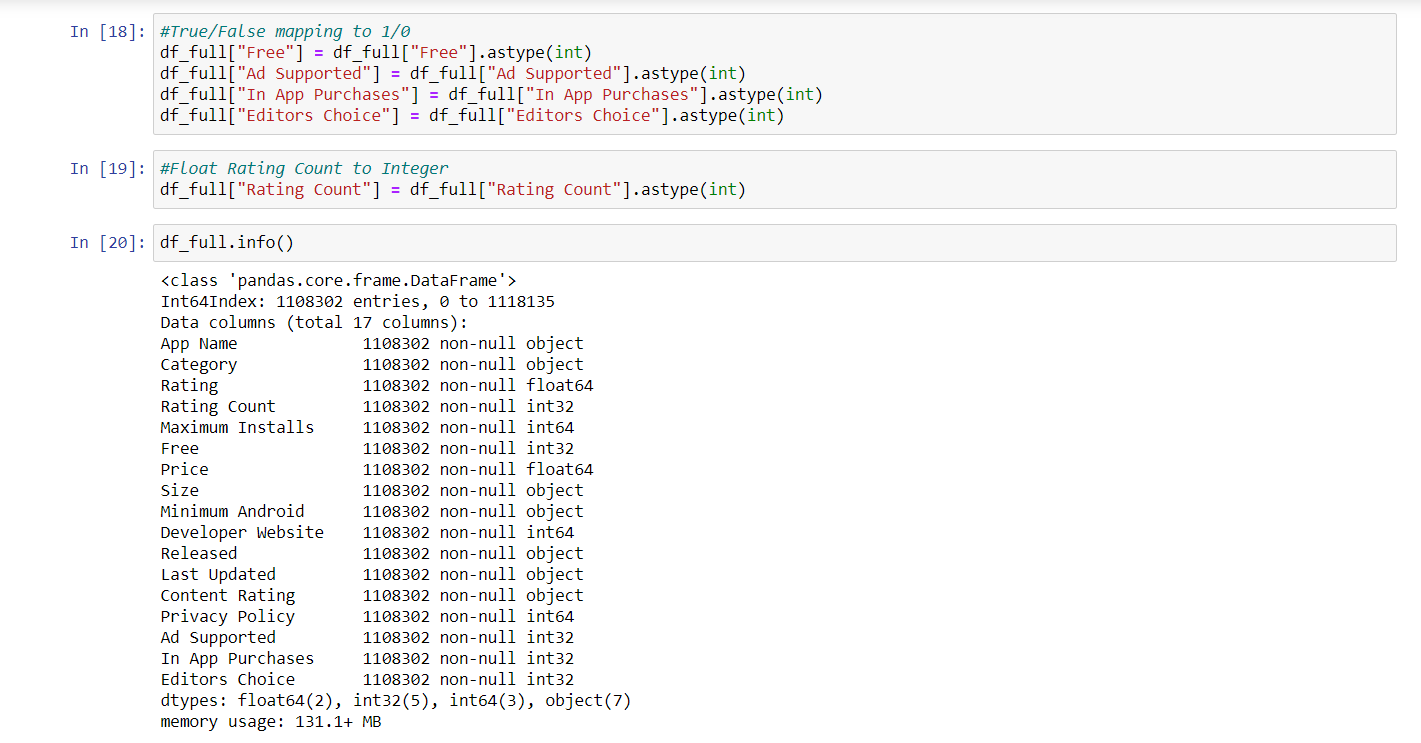
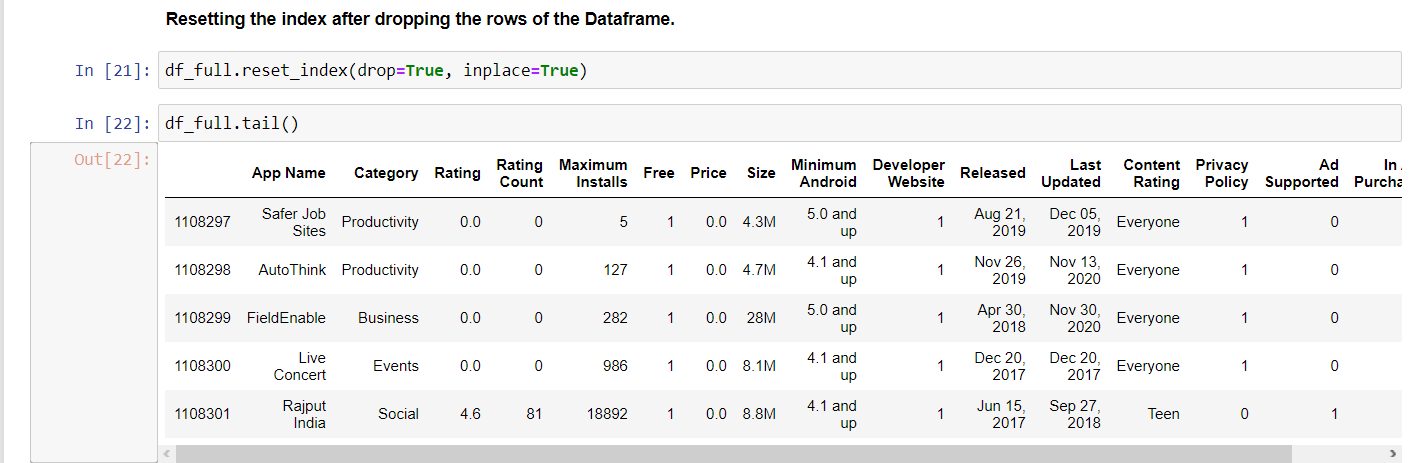
****

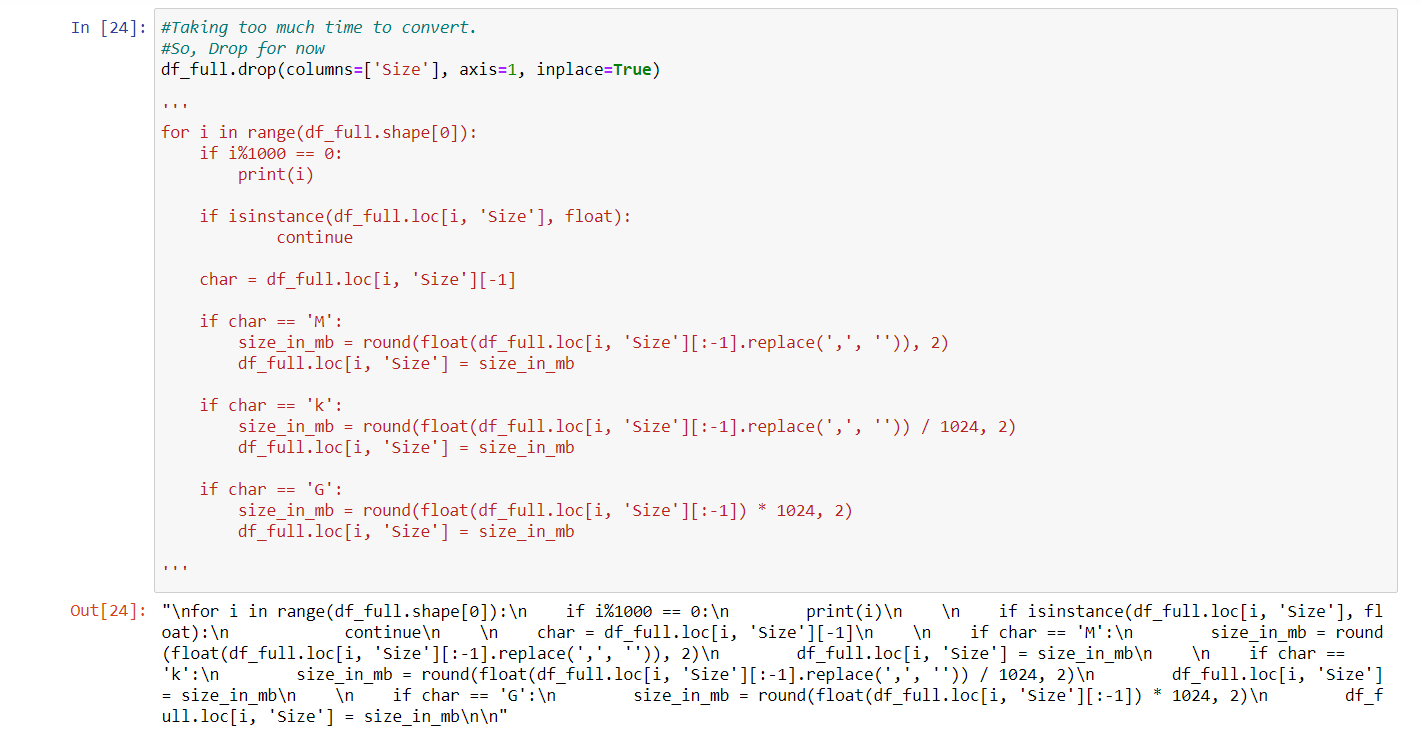
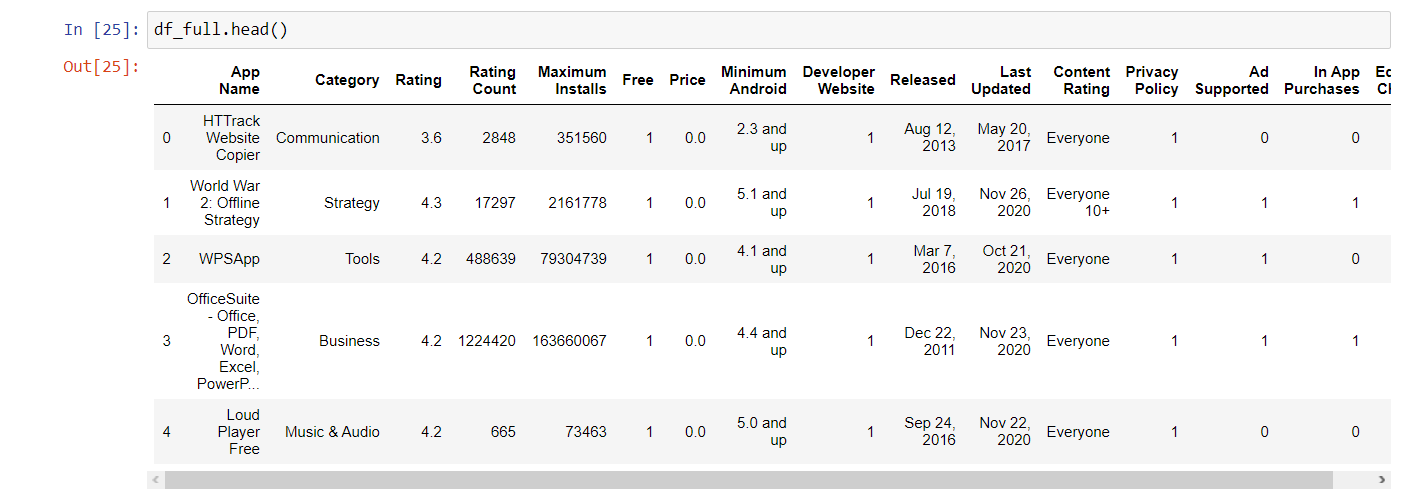
**** ****

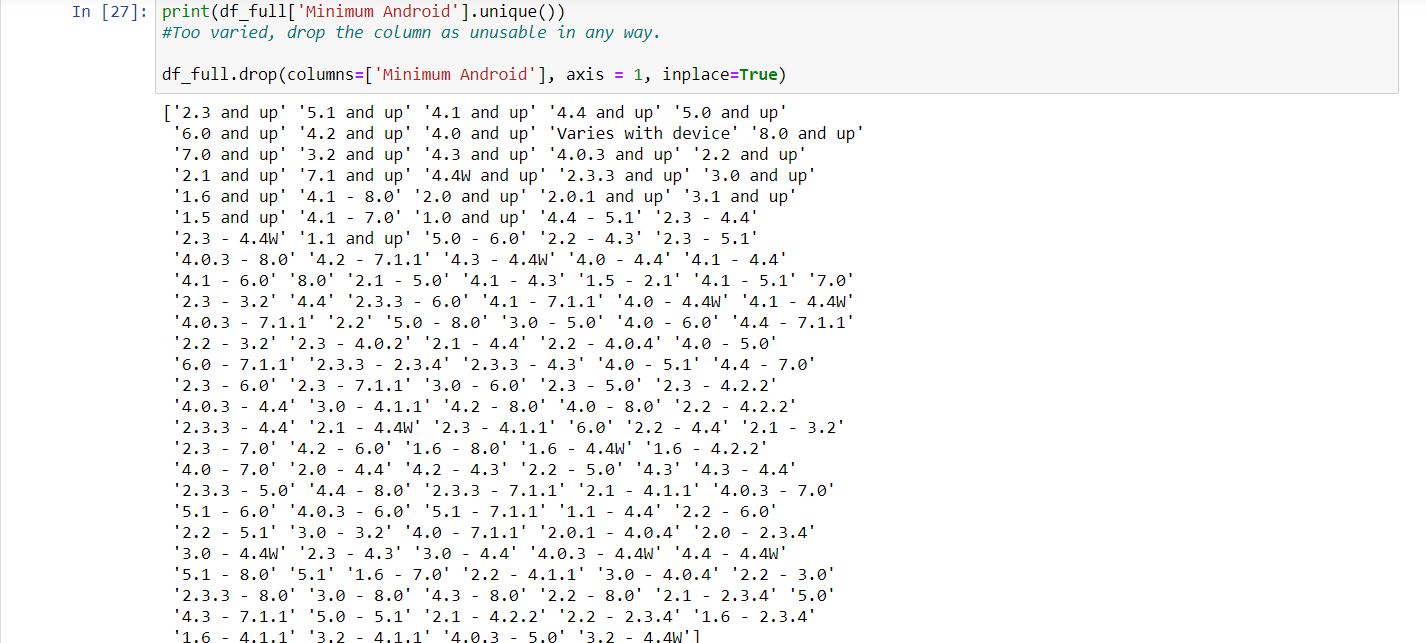
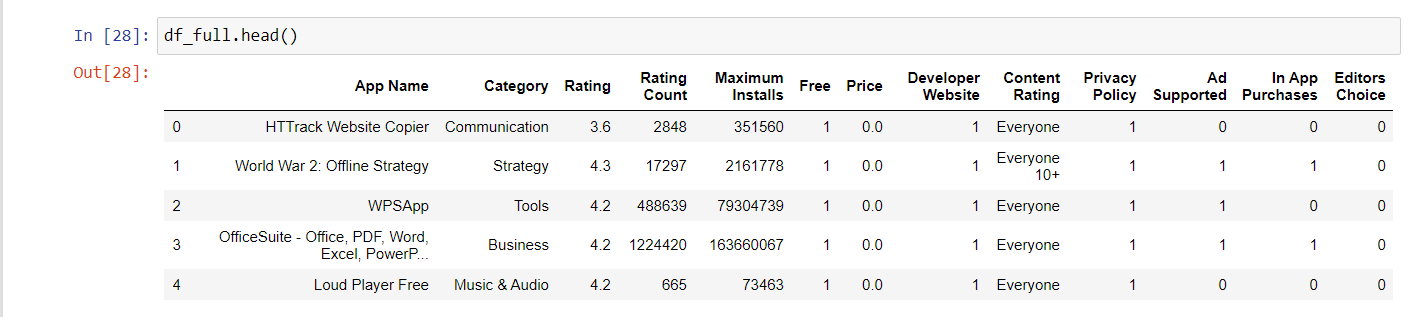
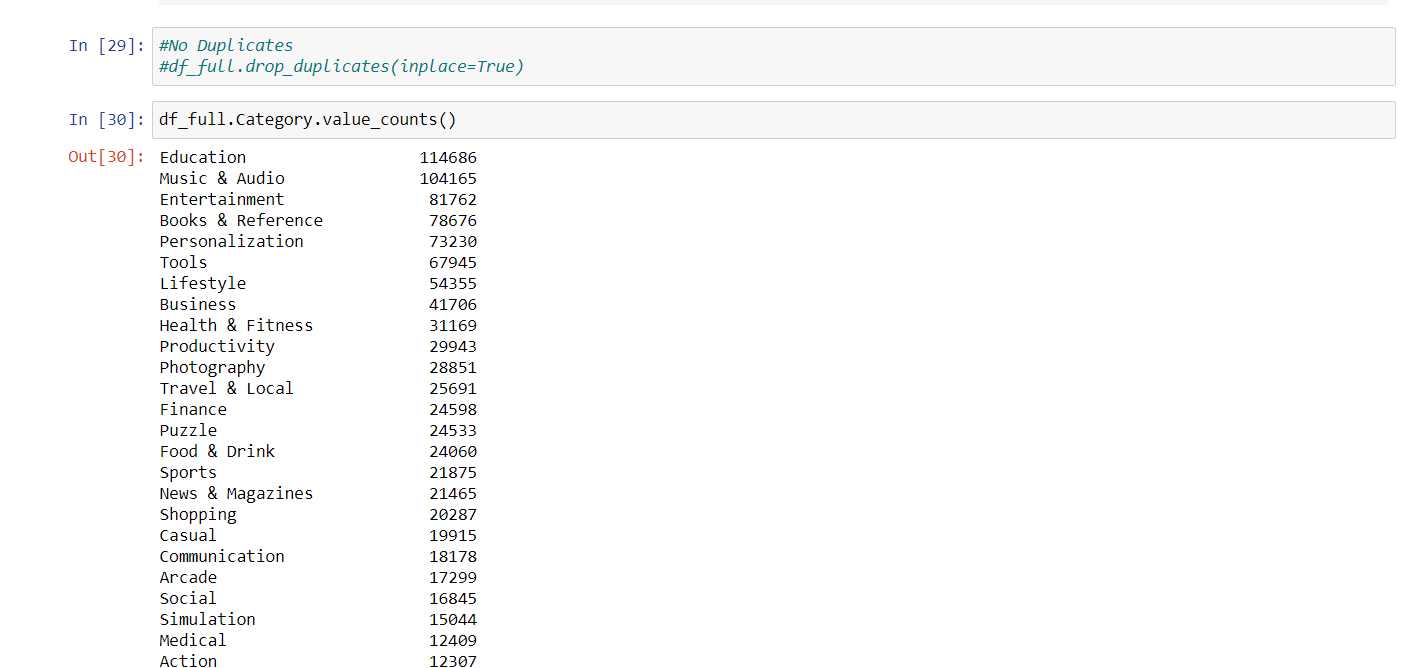
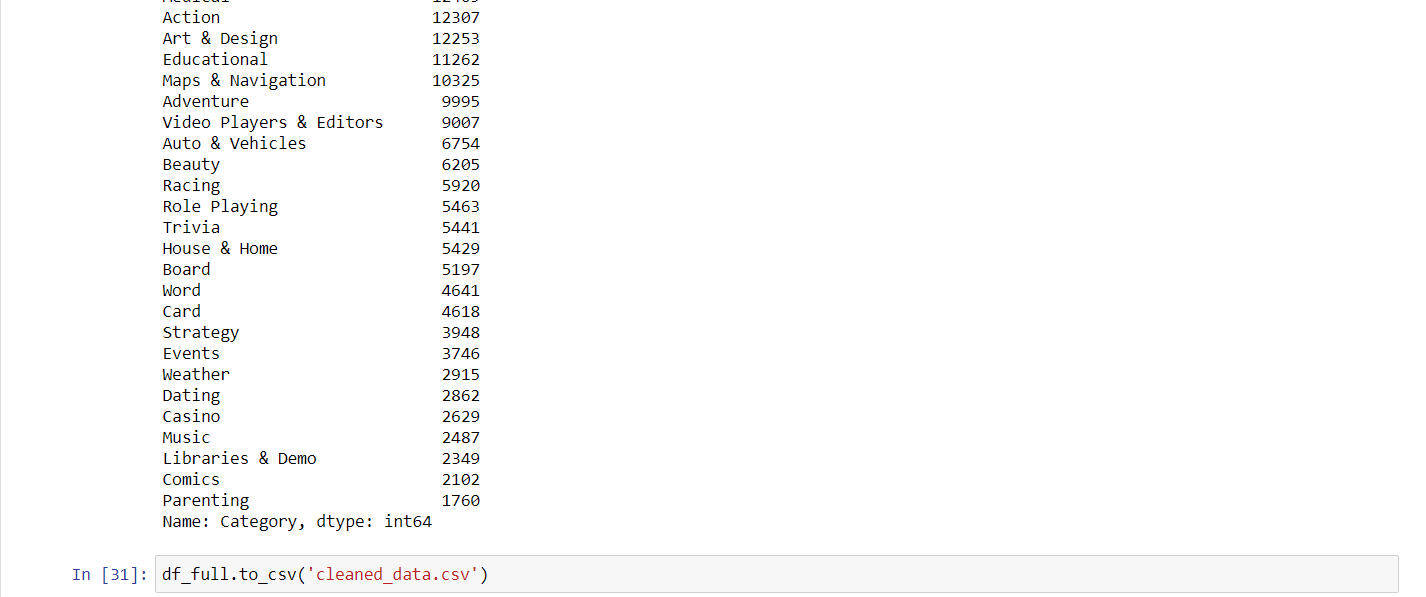
**** ****

**** **** 

**** **** 

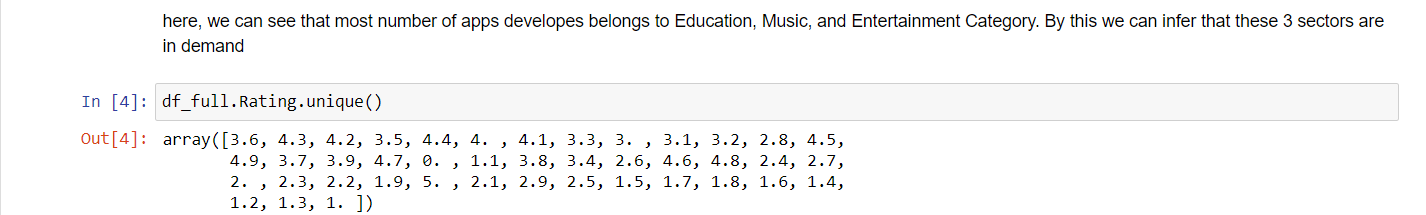
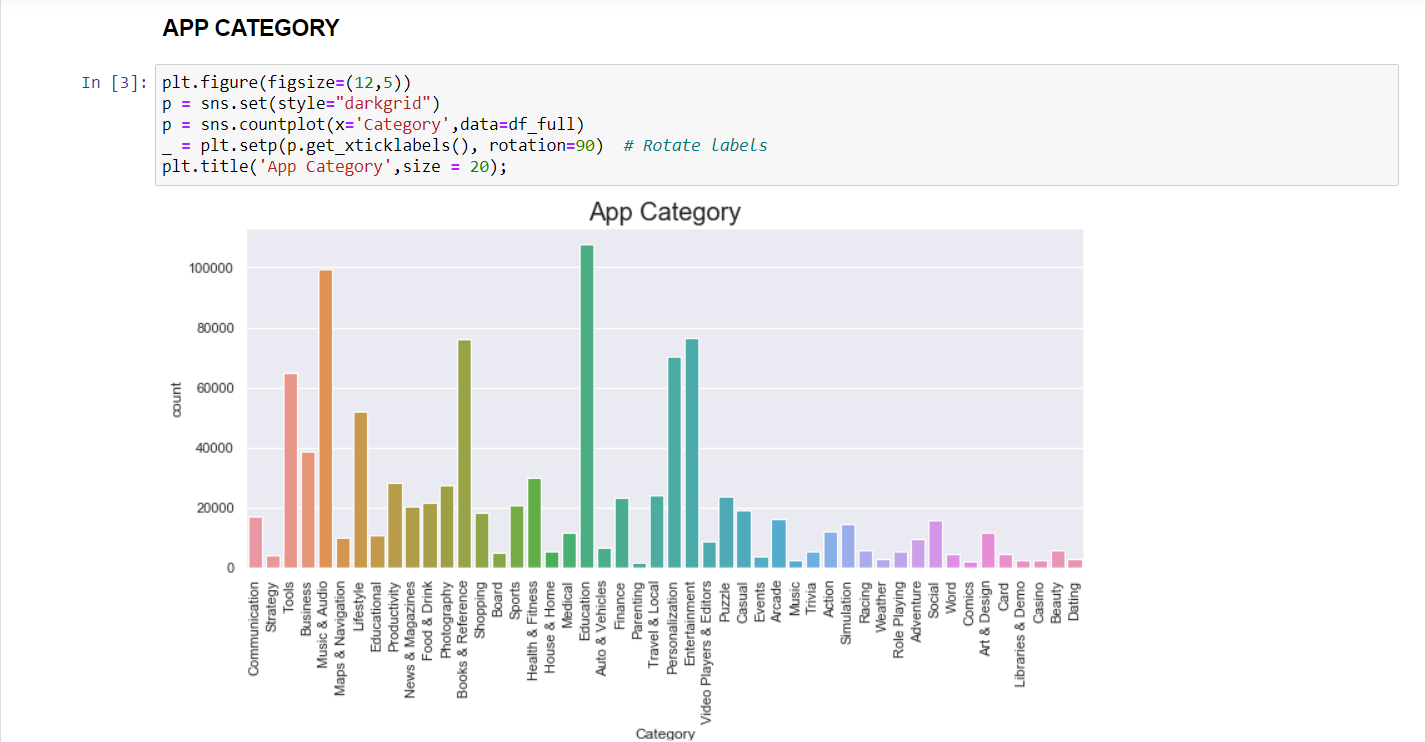
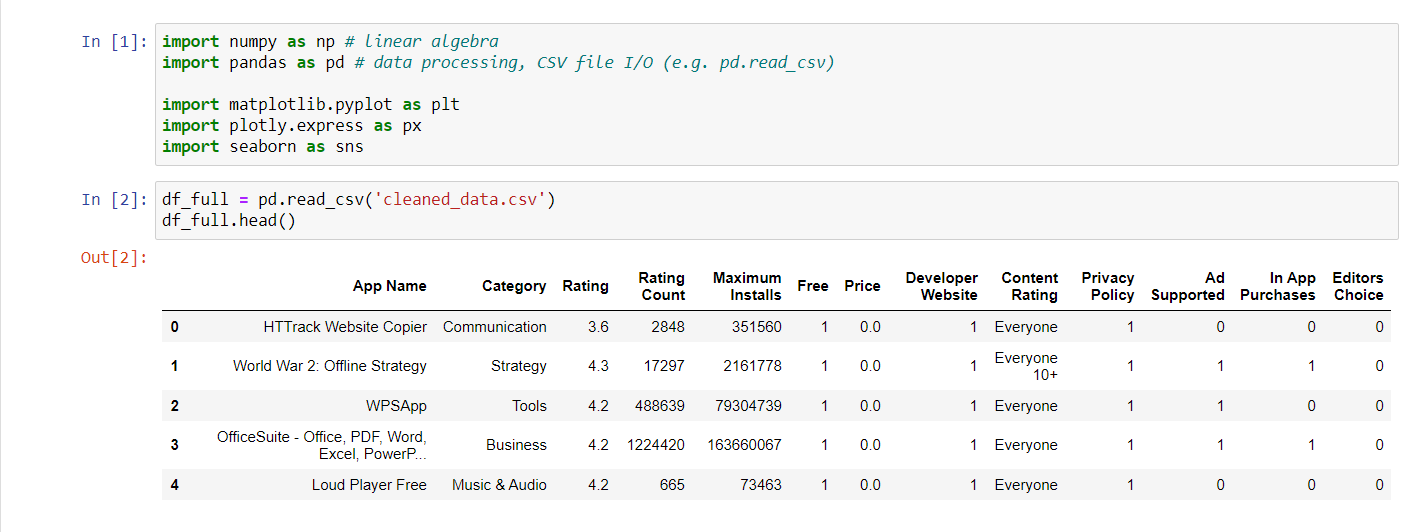
**** **** 

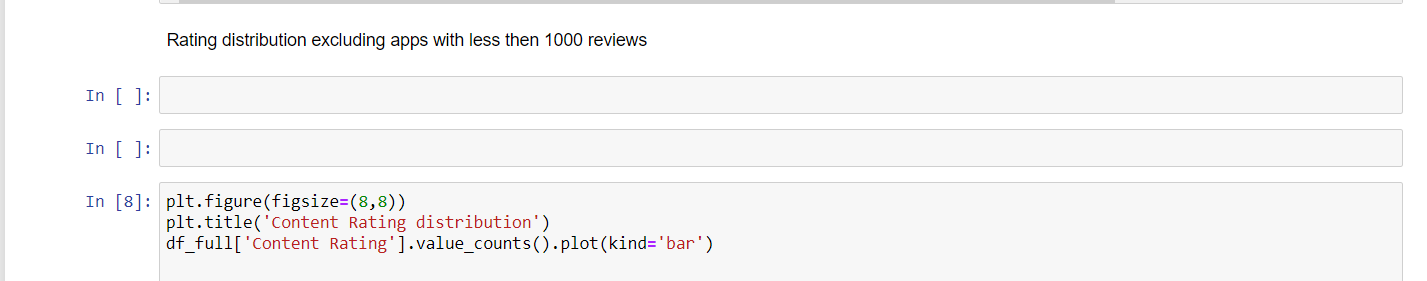
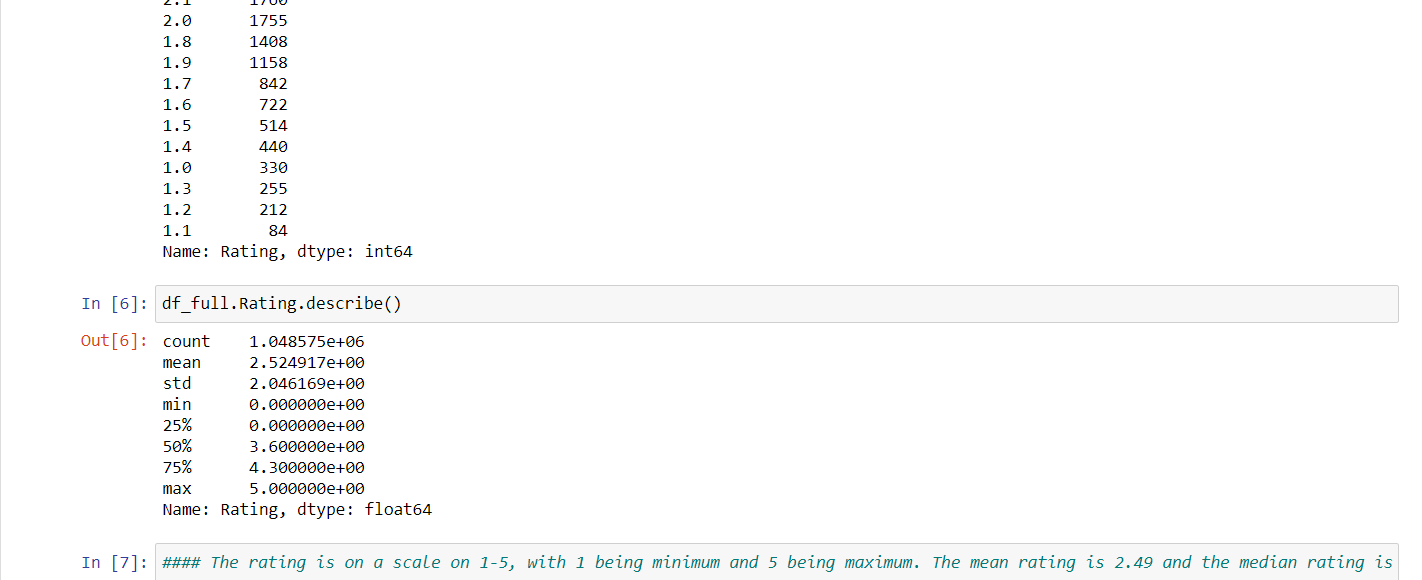
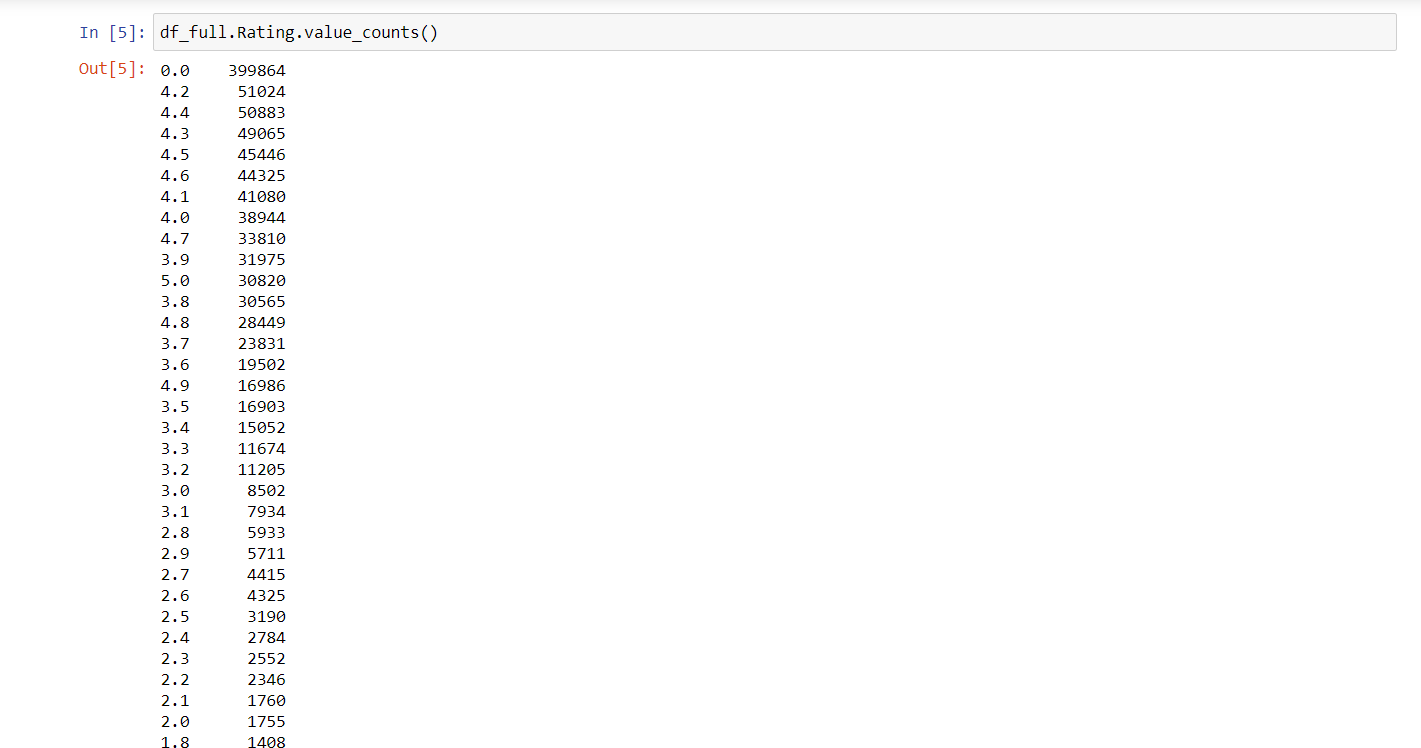
**** **** 

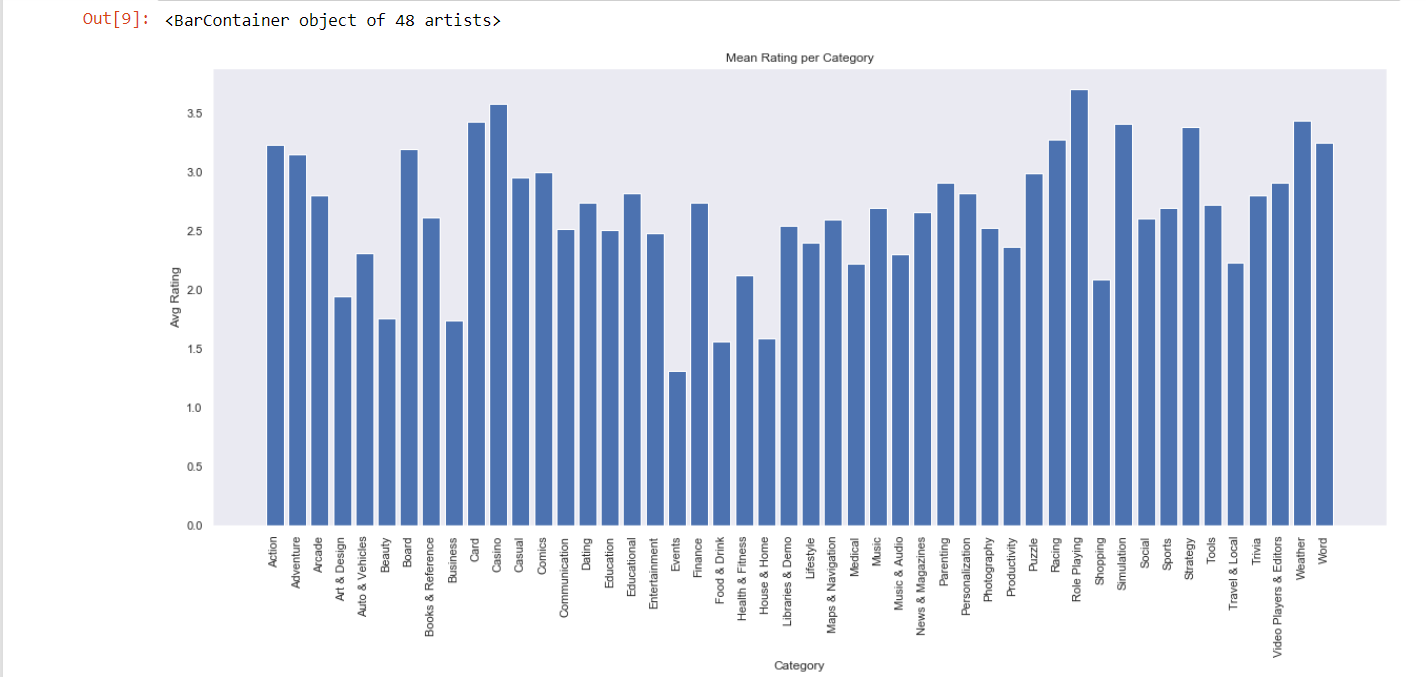
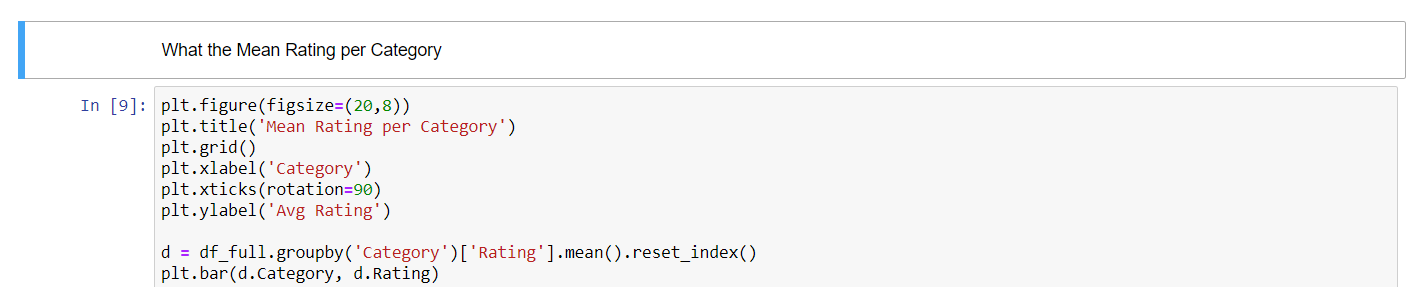
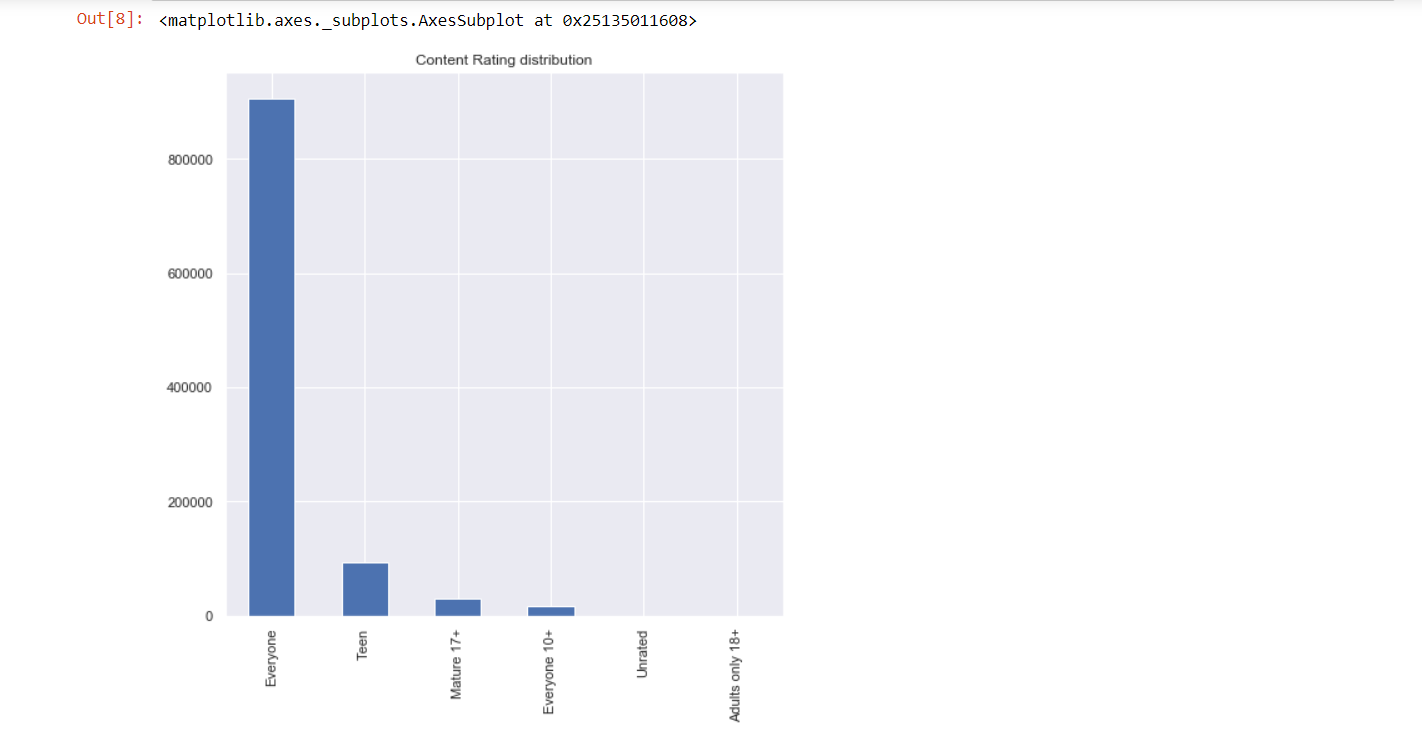
**** ****   

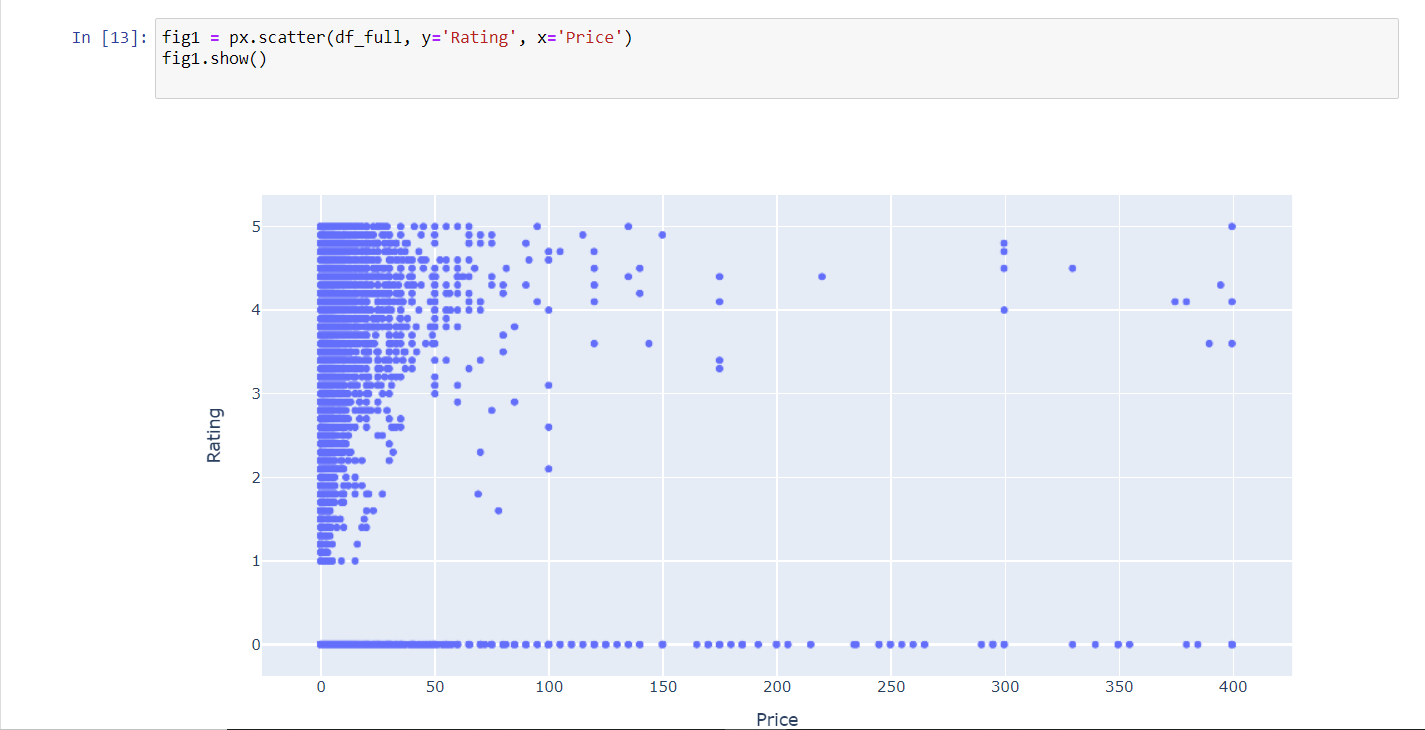
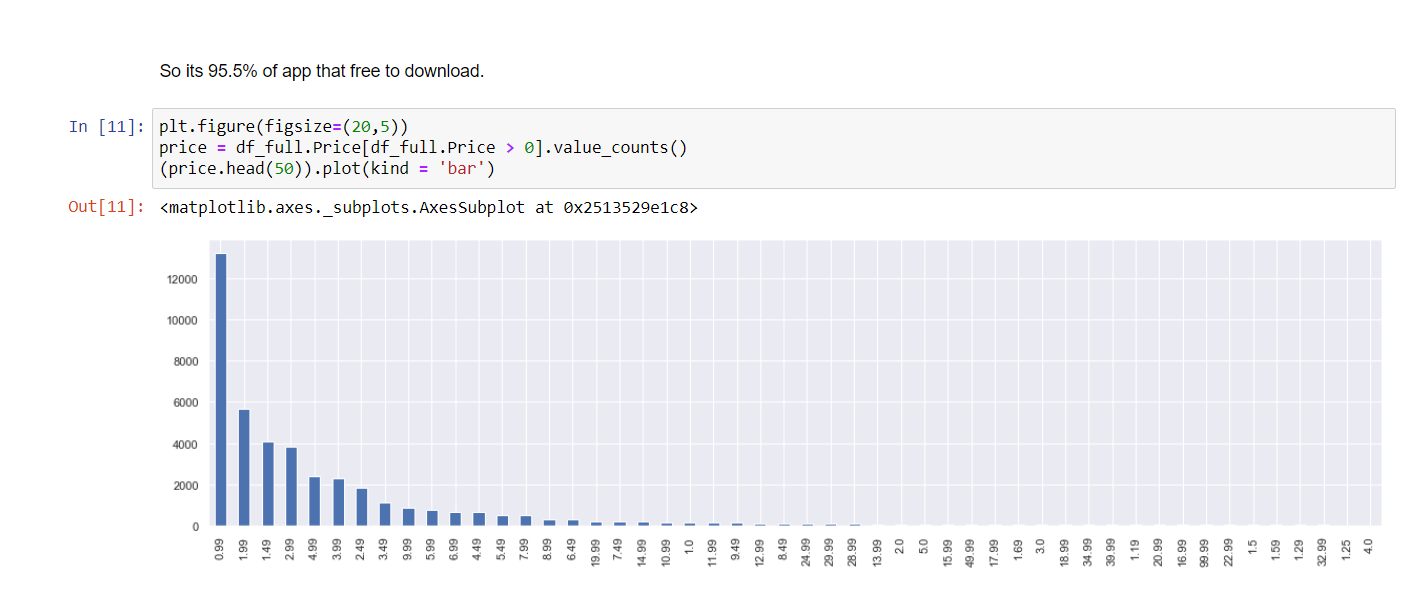
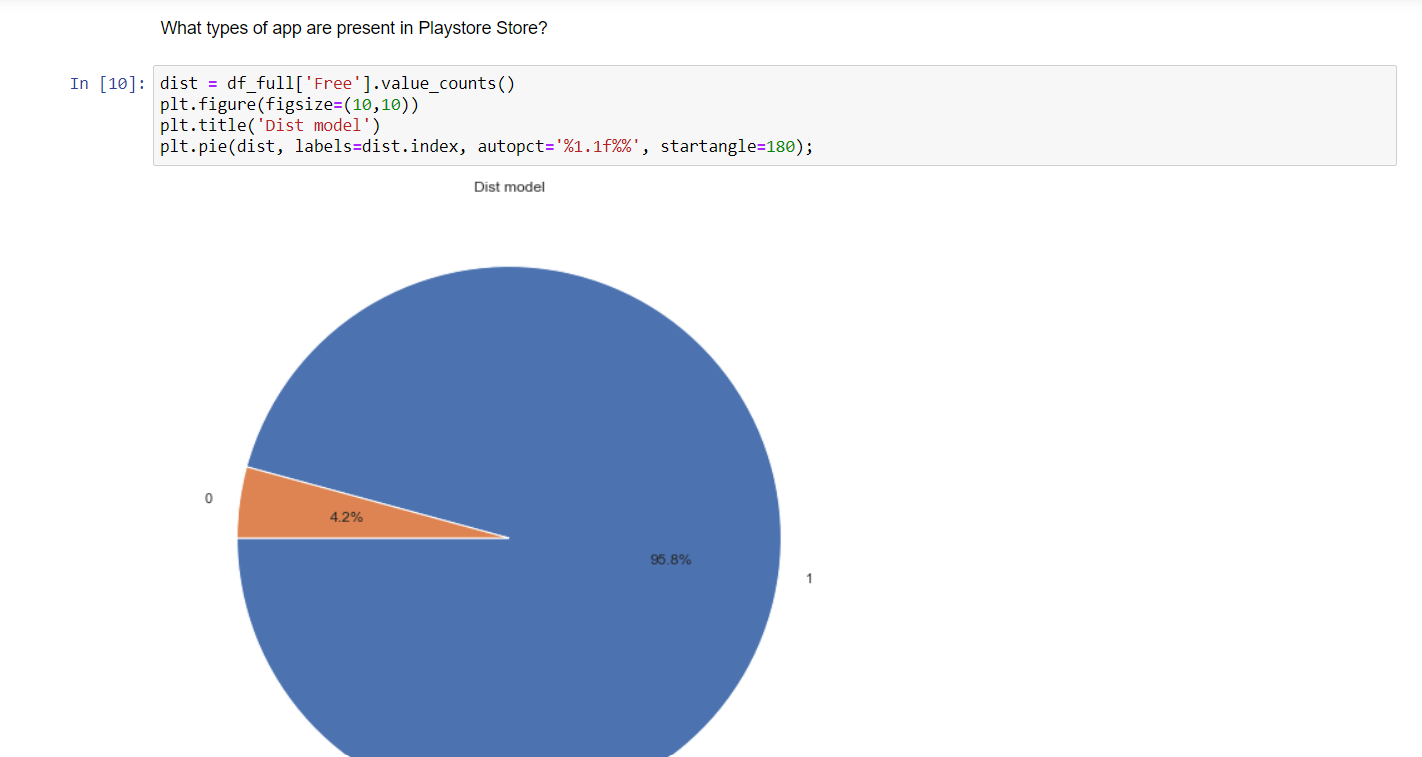
**CODE AND OUTPUT:**

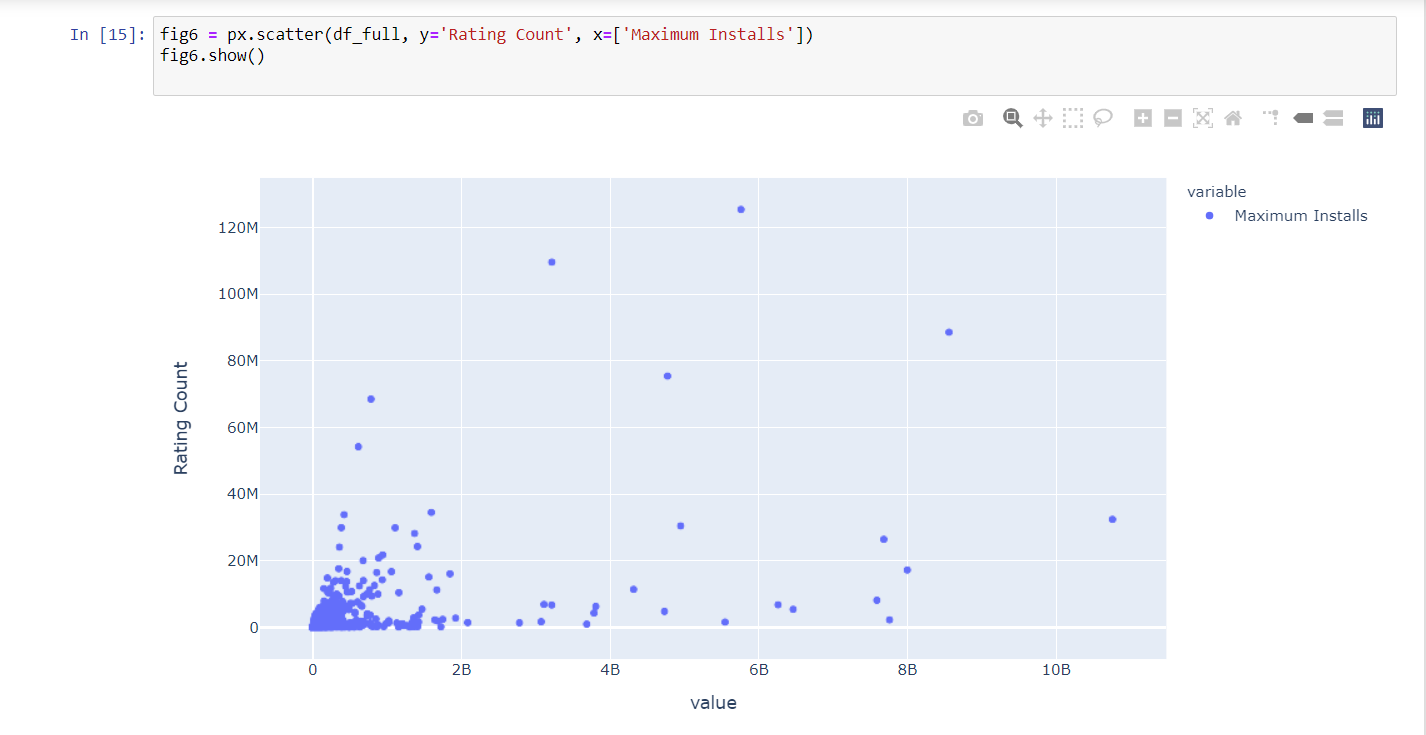
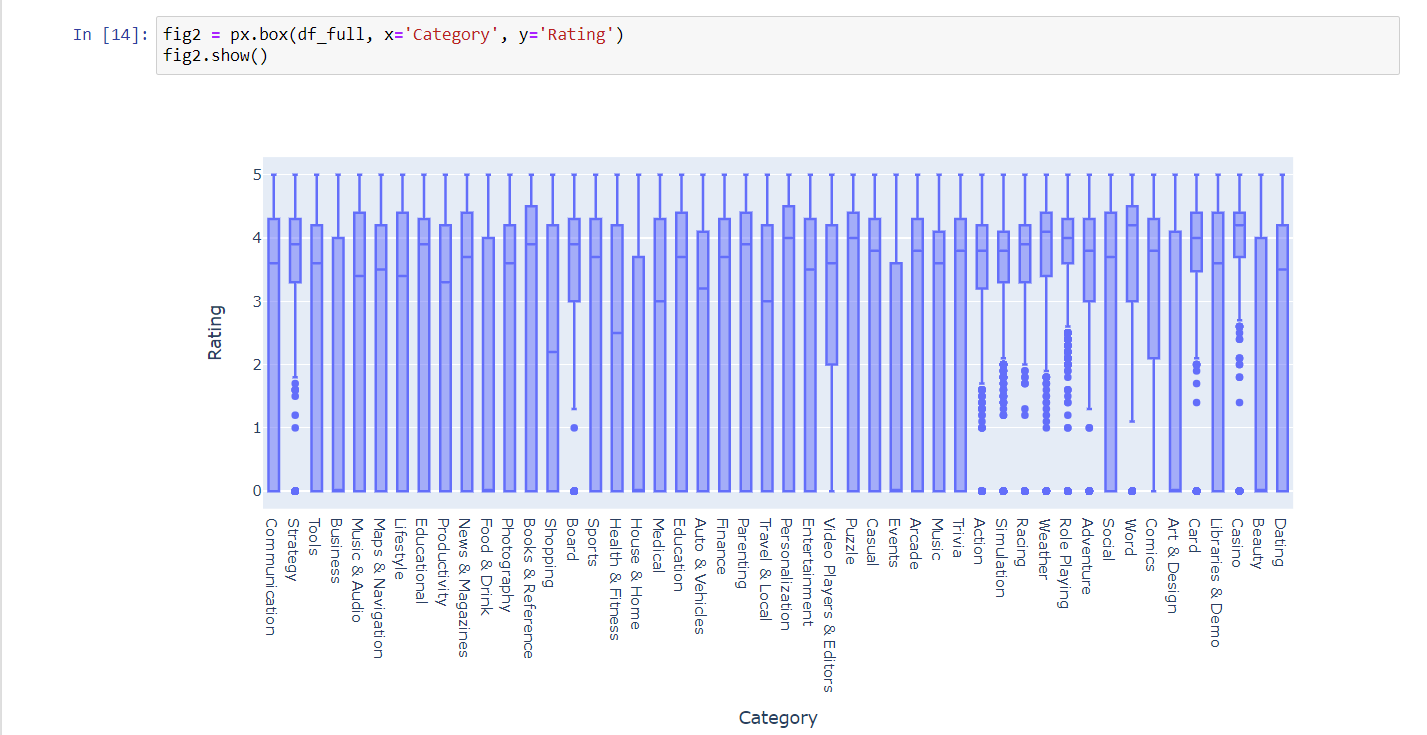
**OBJECTIVE 2:** Visualization of the dataset by using various plots and graphs.

****

****







**CODE AND OUTPUT:**

**OBJECTIVE 3 and 4:**

* Build a model to predict Google Play Store Apps Ratings**.**
* Comparison between different models used for prediction.

