

# Capstone Project

## EDA Analysis on Play Store APP Data Set

### TEAM

Anumalla Rajashekar

Akula Murali Krishna

## OBJECTIVE:

To perform Exploratory Data Analysis on a given dataset

### Exploratory Data Analysis (EDA)

EDA is the process of investigating the dataset to discover patterns, and anomalies (outliers), and form hypotheses based on our understanding of the dataset

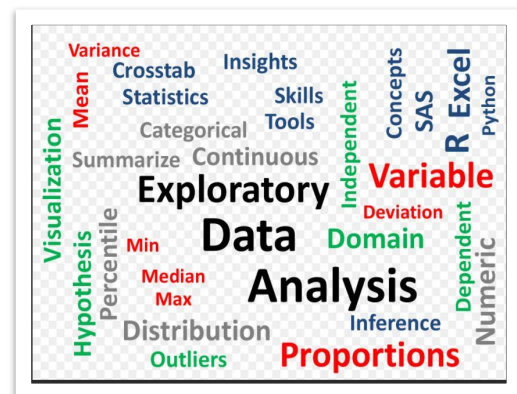
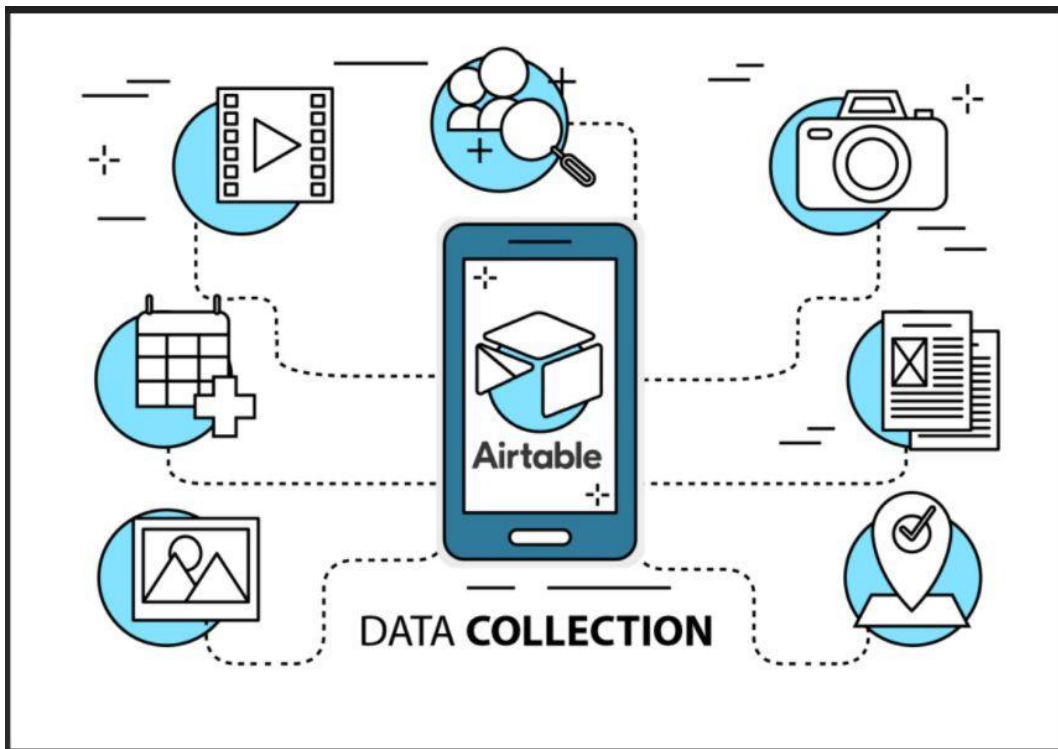


Image source from google

## EDA

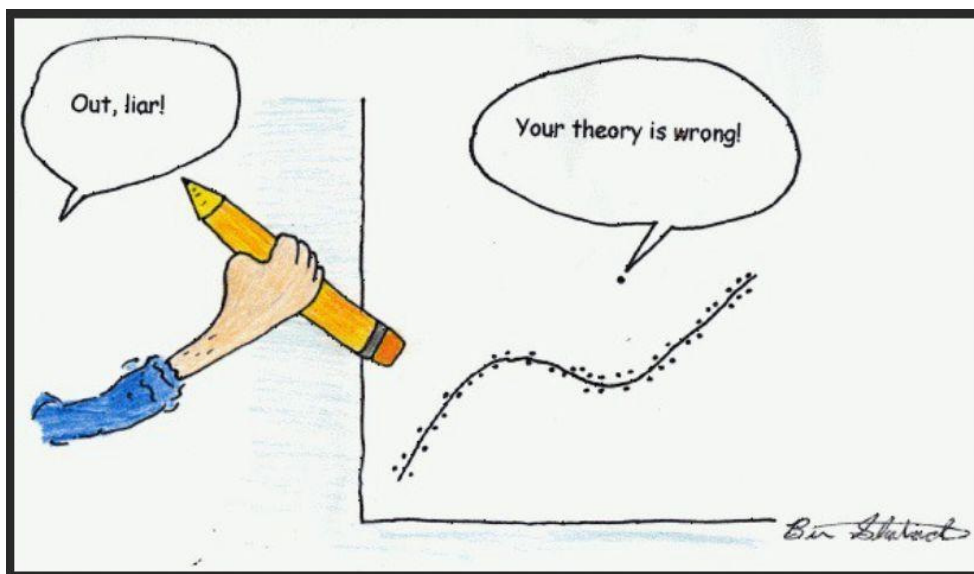
- DATA COLLECTION
- DATA CLEANING
- HYPOTHESIS
- DATA VISUALIZATION



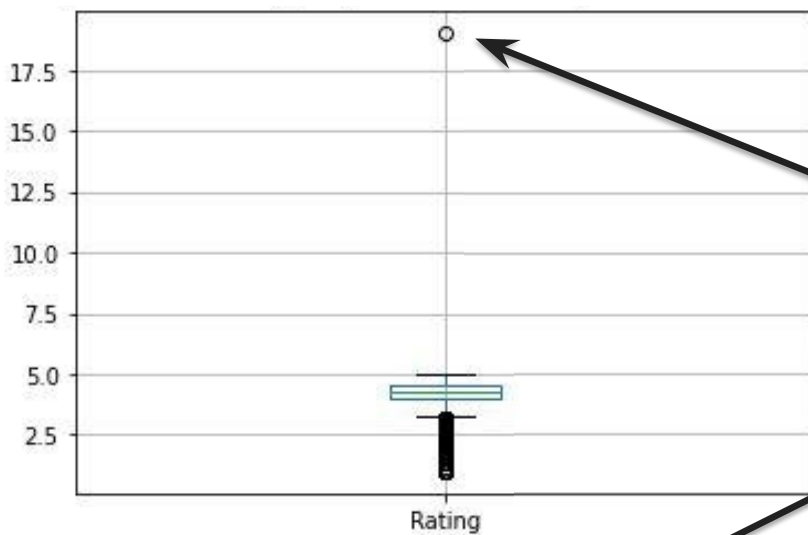
## DATA CLEANING



## DEALING WITH OUTLIER



## OUTLIERS DETECTION



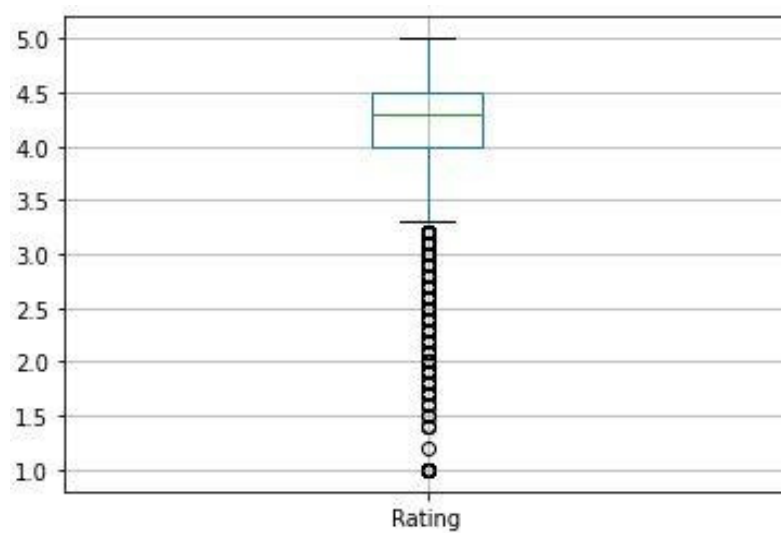
Actions can be perform:

- ✓ Can be corrected (or)
- ✓ Can be dropped

Outlier

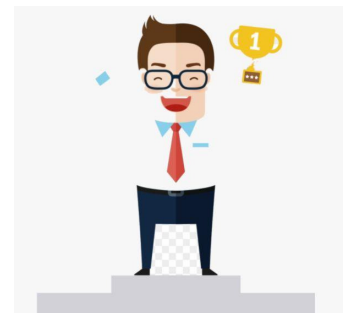
	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver	
10472	Life Made WI-Fi Touchscreen Photo Frame		1.9	19.0	3.0M	1,000+	Free	0	Everyone	NaN	February 11, 2018	1.0.19	4.0 and up	NaN

## AFTER DEALING OUTLIERS



Action took:  
✓ Corrected the data

NO OUTLIERS





## FINDING DUPLICATE RECORDS

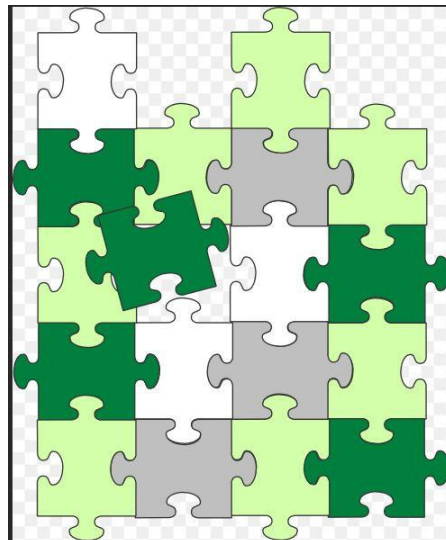


## DEALING WITH MISSING DATA VALUES

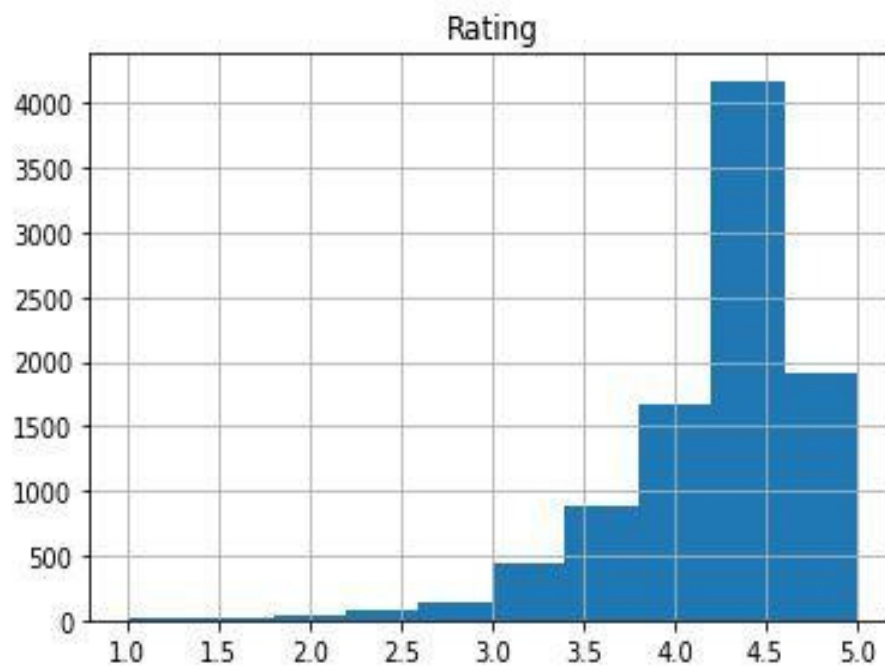


## Detection Of Missing Data

```
App          0
Category     0
Rating       1474
Reviews      0
Size         0
Installs     0
Type         1
Price        0
Content Rating 0
Genres       1
Last Updated 0
Current Ver  8
Android Ver  2
dtype: int64
```



## Rating v/s Number of Apps



Left Side Skewed



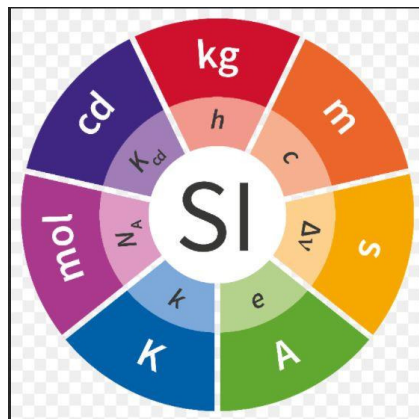
## Handling Of Missing Data



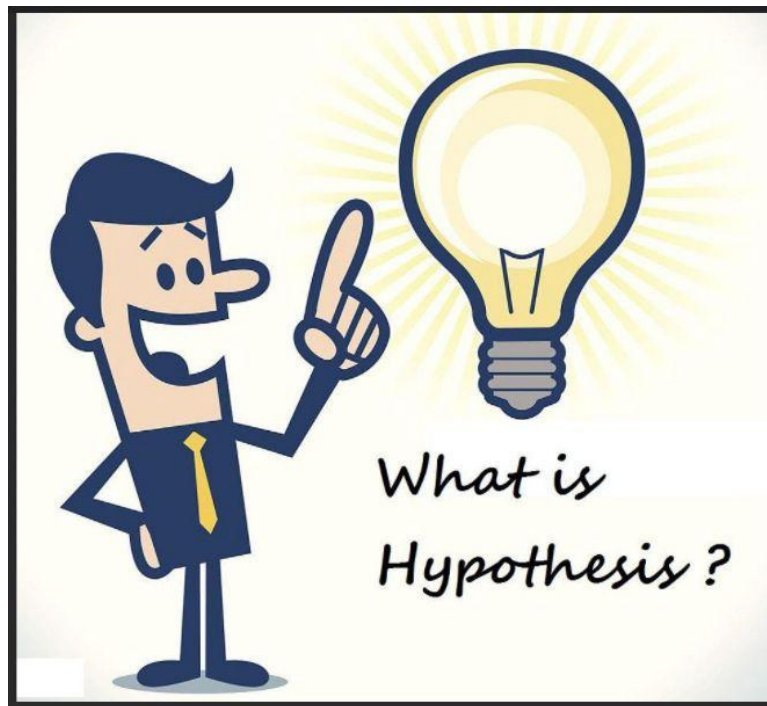
```
App          0
Category     0
Rating       0
Reviews      0
Size         0
Installs     0
Type         0
Price        0
Content Rating 0
Genres       0
Last Updated 0
Current Ver  0
Android Ver  0
dtype: int64
```

# Data Formatting

- Units
- Symbols
- AlphaNumeric

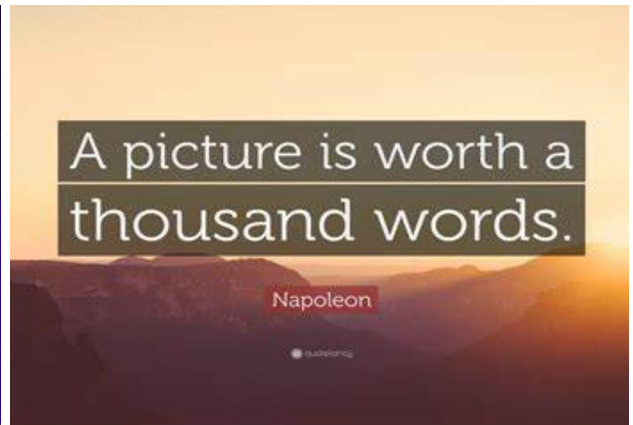
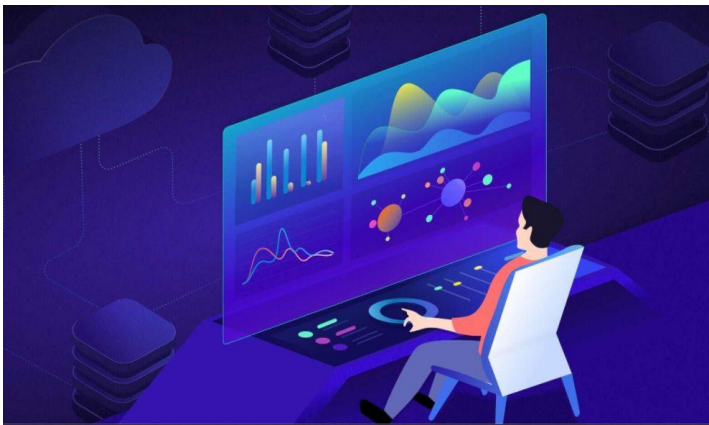


€ EU: Euro	\$ US: Dollar	¥ China: Yuan	¥ Japan: Yen
DM Germany: Deutsche Mark	₣ France: Franc	£ UK: Pound	₤ Italy: Lira
R\$ Brazil: Real	C\$ Canada: Dollar	₽ Russia: Ruble	₹ India: Rupee
₧ Spain: Peseta	A\$ Australia: Dollar	Mex\$ Mexico: Peso	₩ South Korea: Won
f Netherlands: Guilder	TL Turkey: Lira	Rp Indonesia: Rupiah	SFr Switzerland: Franc



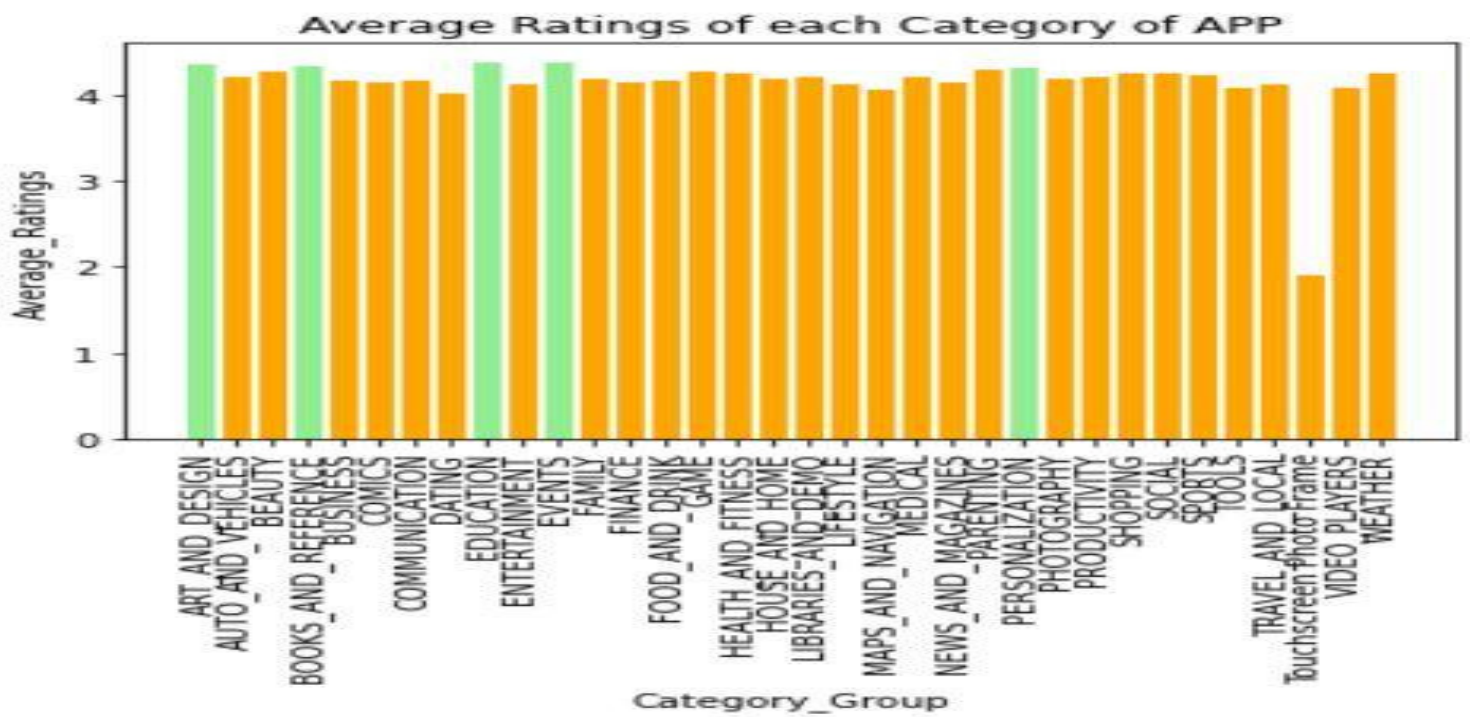
## DATA VISUALIZATION

AI

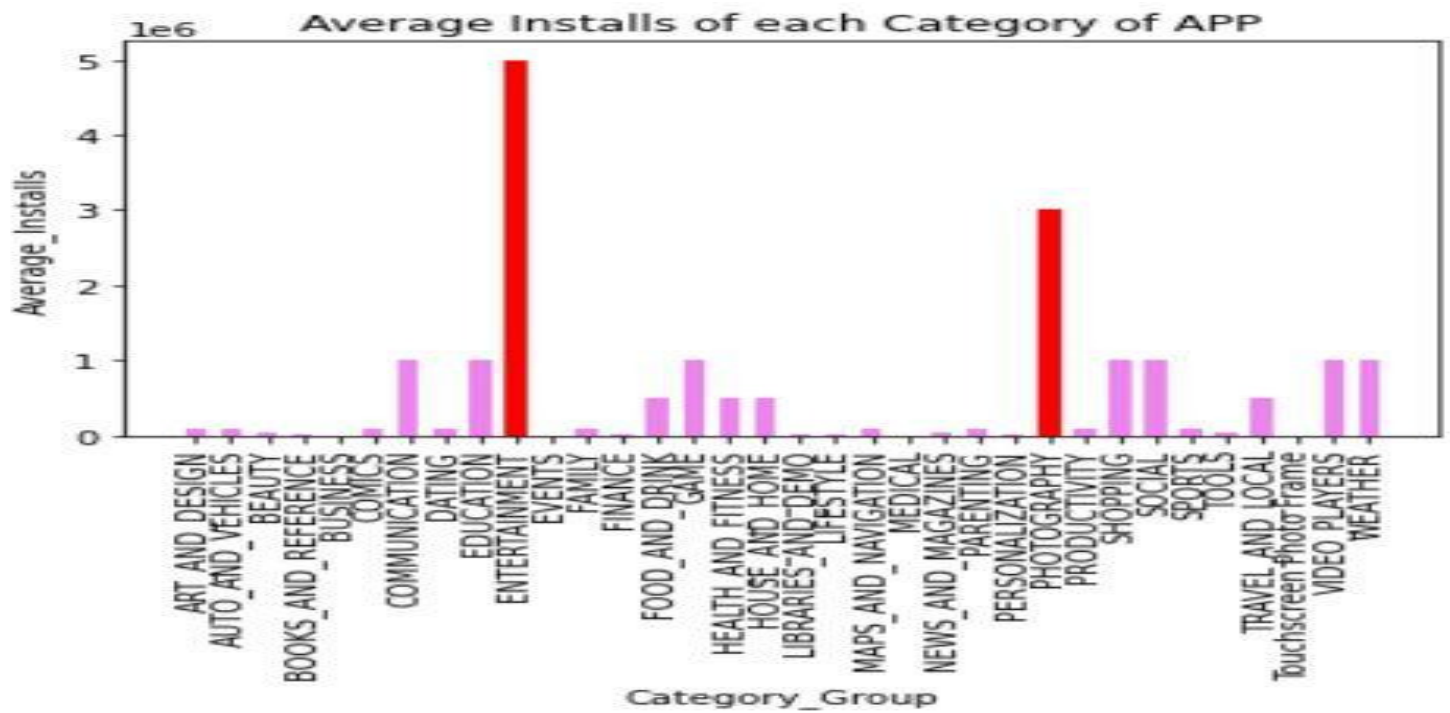




## 1.Categories group vs Average Ratings

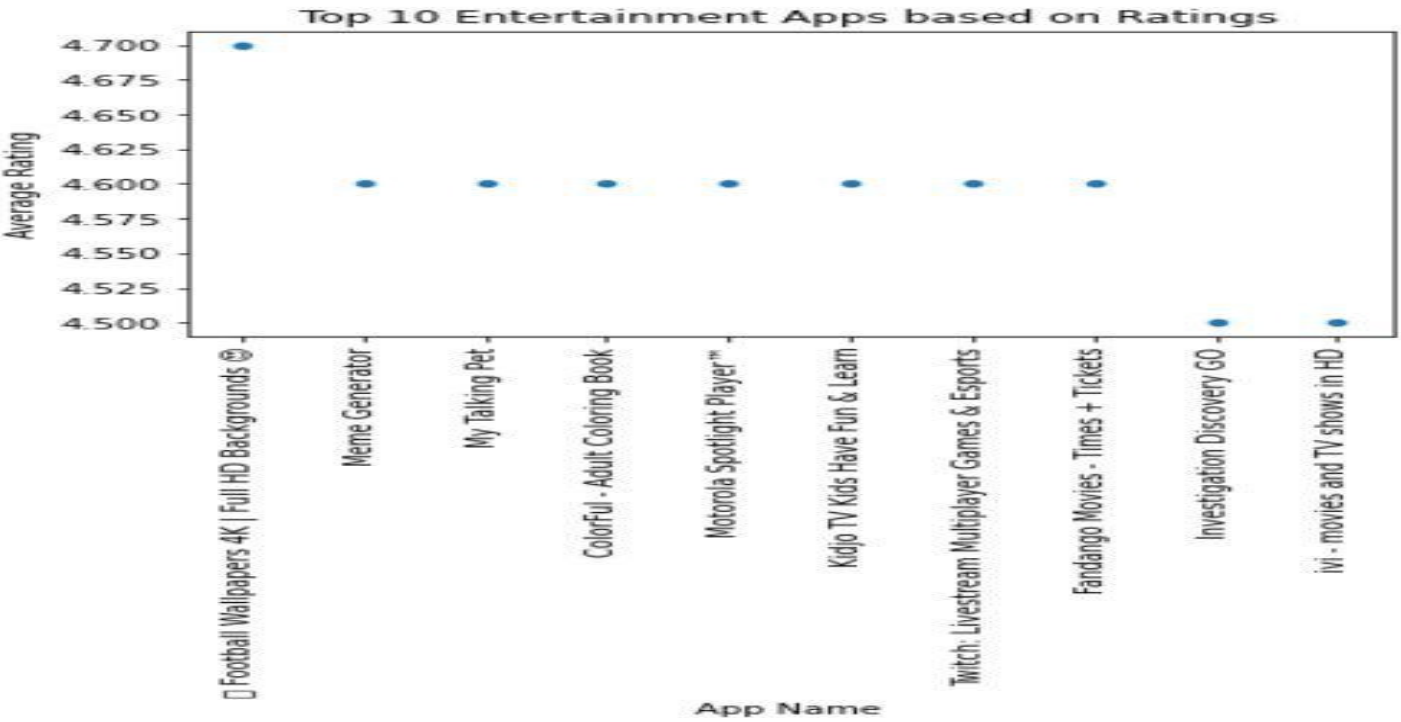


## 2.Categories group vs Average Install



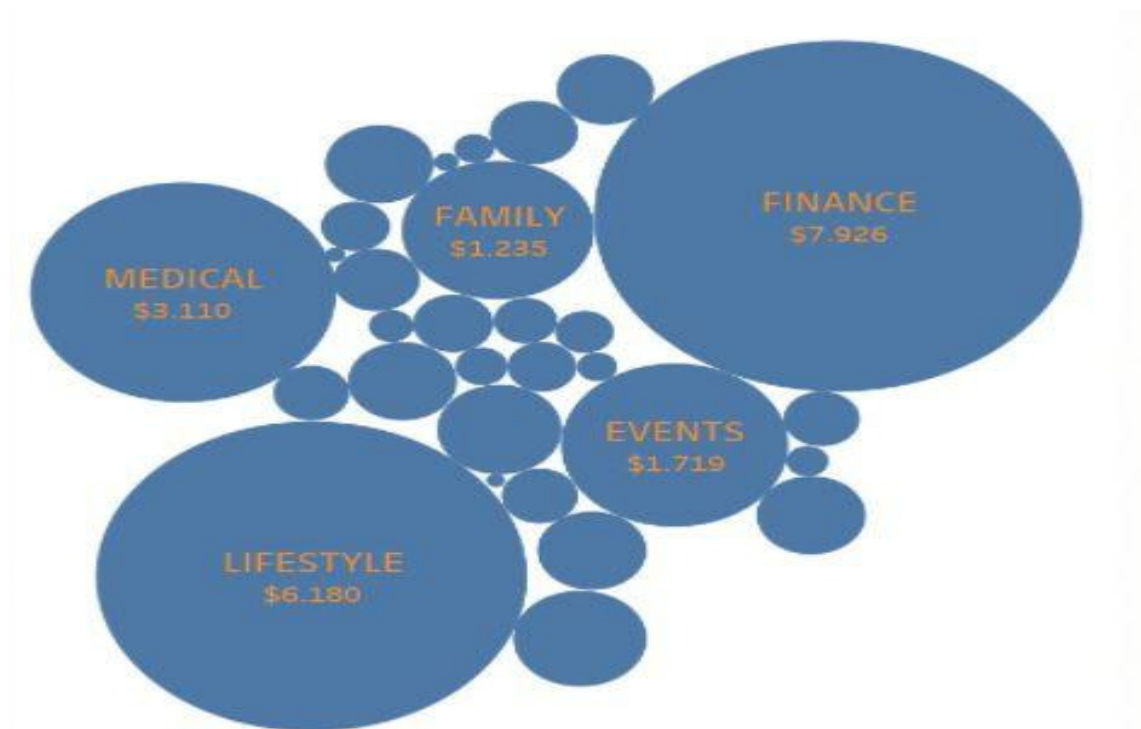
3.Average Rating Vs App

AI



## 4.Categories Based On Average Price

AI



## CONCLUSION

We believe that we have conclude our objective on this particular dataset.

### Overview of tasks performed:

- ☐ Data Cleaning
- ☐ Hypotheses Formation
- ☐ Data Visualization to support hypotheses

**Thank You**

