Google Play Apps Rating Analysis

- T5 Bootcamp Data Science Project – Maram Hussain

Project Goal

Identify the apps that are going to be good for Google to promote by analyzing the App ratings using machine learning models.

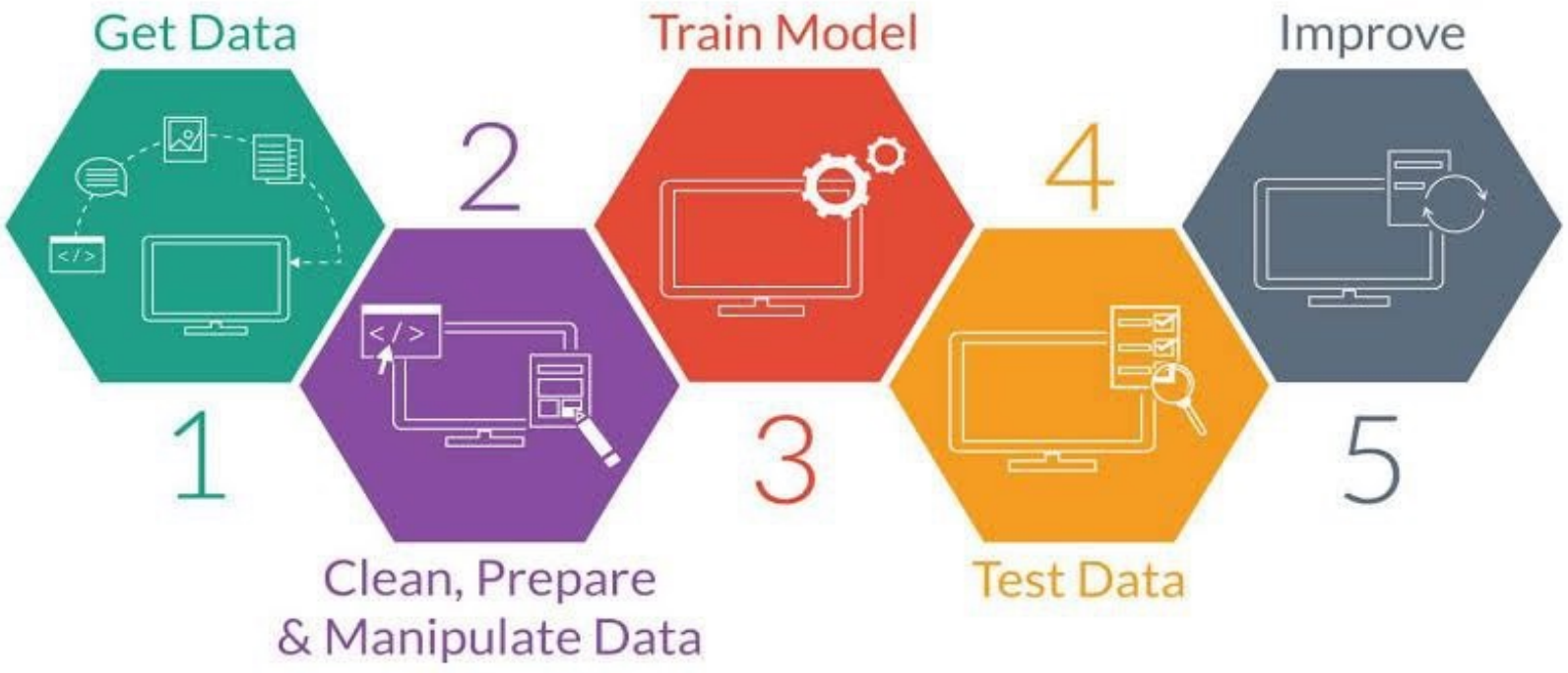
Google Play Apps

The dataset is provided in .csv format from Kaggle. It contains 10842 apps, each app has 13 features

Dataset Sample

```
data.head()
```

	Rating	App	Category	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
0	4.1	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	159	19M	10,000+	Free	0	Everyone	Art & Design	07-Jan-18	1.0.0	4.0.3 and up
1	3.9	Coloring book moana	ART_AND_DESIGN	967	14M	500,000+	Free	0	Everyone	Art & Design;Pretend Play	15-Jan-18	2.0.0	4.0.3 and up
2	4.7	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	87510	8.7M	5,000,000+	Free	0	Everyone	Art & Design	01-Aug-18	1.2.4	4.0.3 and up
3	4.5	Sketch - Draw & Paint	ART_AND_DESIGN	215644	25M	50,000,000+	Free	0	Teen	Art & Design	08-Jun-18	Varies with device	4.2 and up
4	4.3	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	967	2.8M	100,000+	Free	0	Everyone	Art & Design;Creativity	20-Jun-18	1.1	4.4 and up



The main steps of machine learning models

Dataset after pre-processing

```
data.head()
```

	Rating	App	Category	Reviews	Size	Installs	Type	Price	Content Rating
0	4.1	6962	0	159.0	19.0	10000.0	0	0.0	1
1	3.9	2632	0	967.0	14.0	500000.0	0	0.0	1
2	4.7	8656	0	87510.0	8.7	5000000.0	0	0.0	1
3	4.5	7827	0	215644.0	25.0	50000000.0	0	0.0	4
4	4.3	7022	0	967.0	2.8	100000.0	0	0.0	1

Modeling

The random forest regressor model was used to predict the rating. The official metric was the R^2 score of the model, where the model tested on the R^2 score , Root Mean Squared Error(RMSE), Mean Squared Error(MSE) and Mean Absolute Error(MAE).

The result:

R_2=85%

MSE=4%

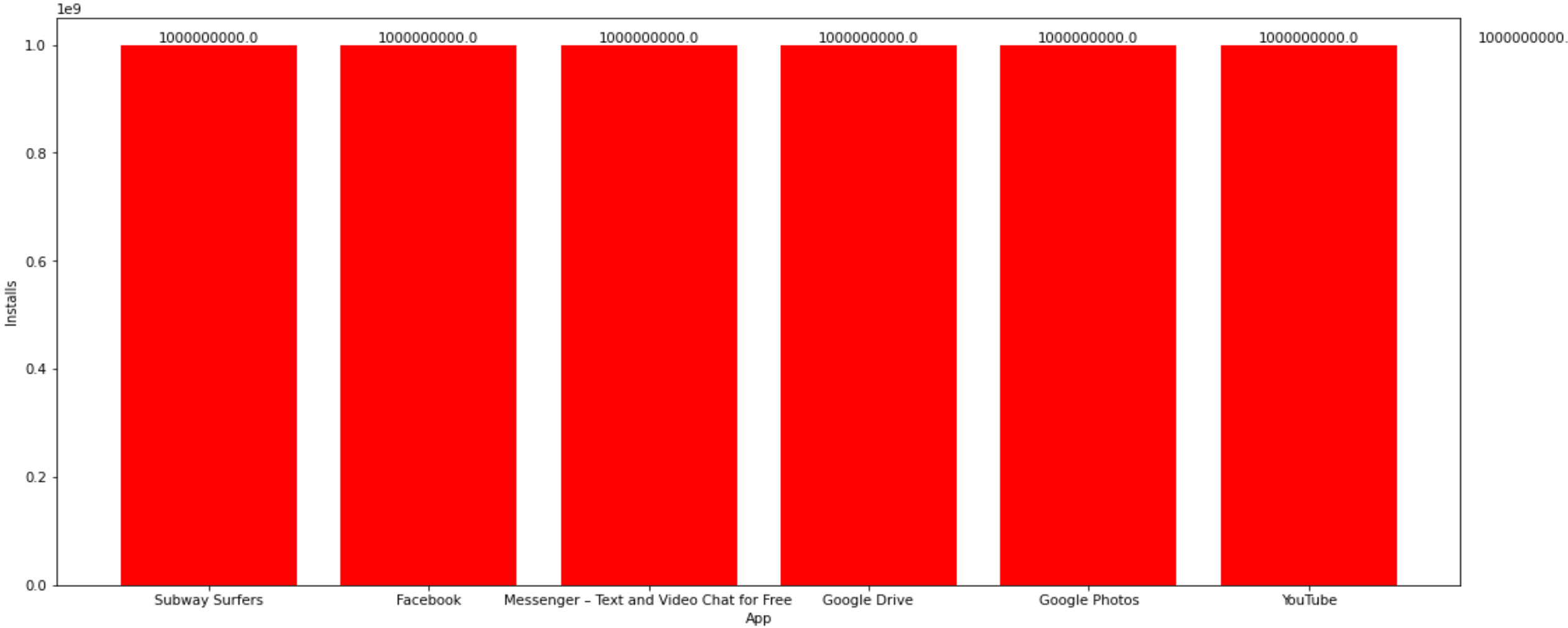
RMSE=19%

MAE=12%

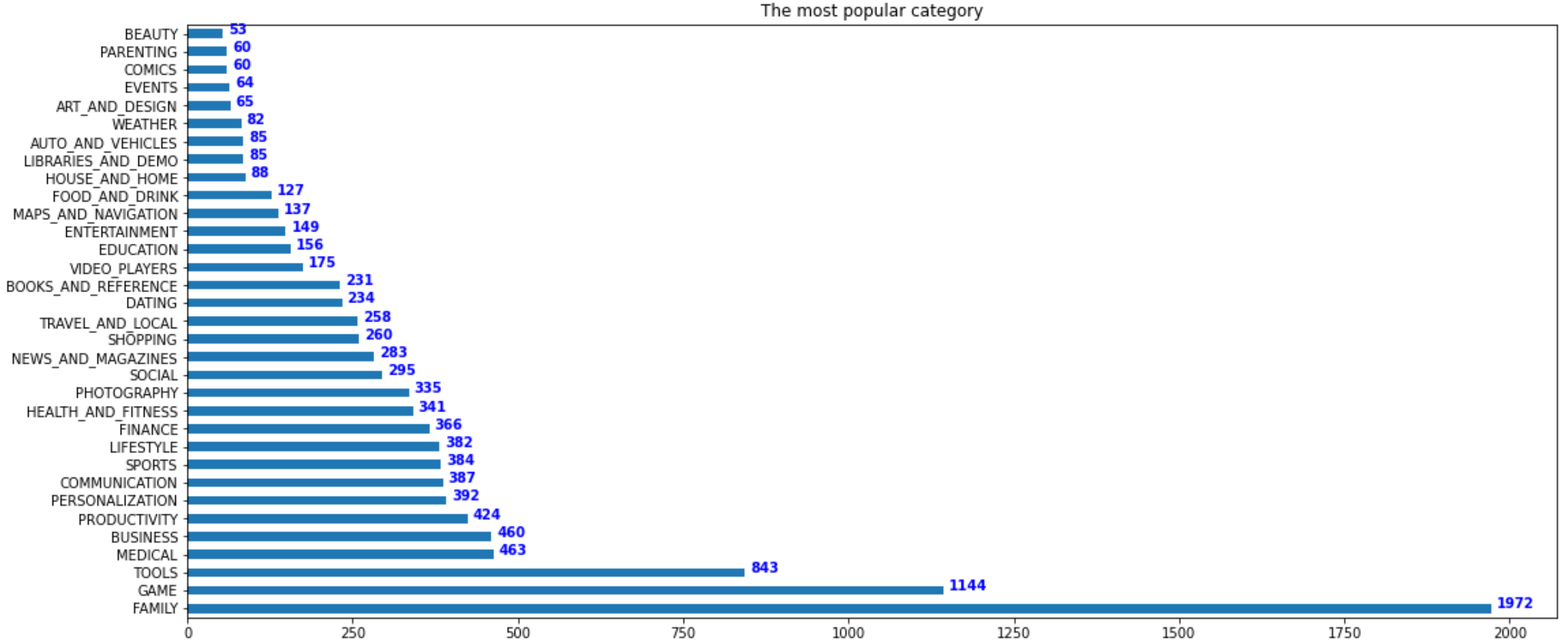
The result of random forest regressor:

	R_2_train	R_2_test	MSE_train	MSE_test	RMSE_train	RMSE_test	MAE_train	MAE_test
LinearRegression	0.0109182	0.0113057	0.260419	0.270809	0.510312	0.520394	0.355672	0.368513
RandomForestRegressor	0.850832	0.201294	0.0392749	0.21877	0.198179	0.467729	0.122262	0.309905
GradientBoostingRegressor	0.310721	0.183814	0.181483	0.223558	0.426008	0.47282	0.287826	0.325634

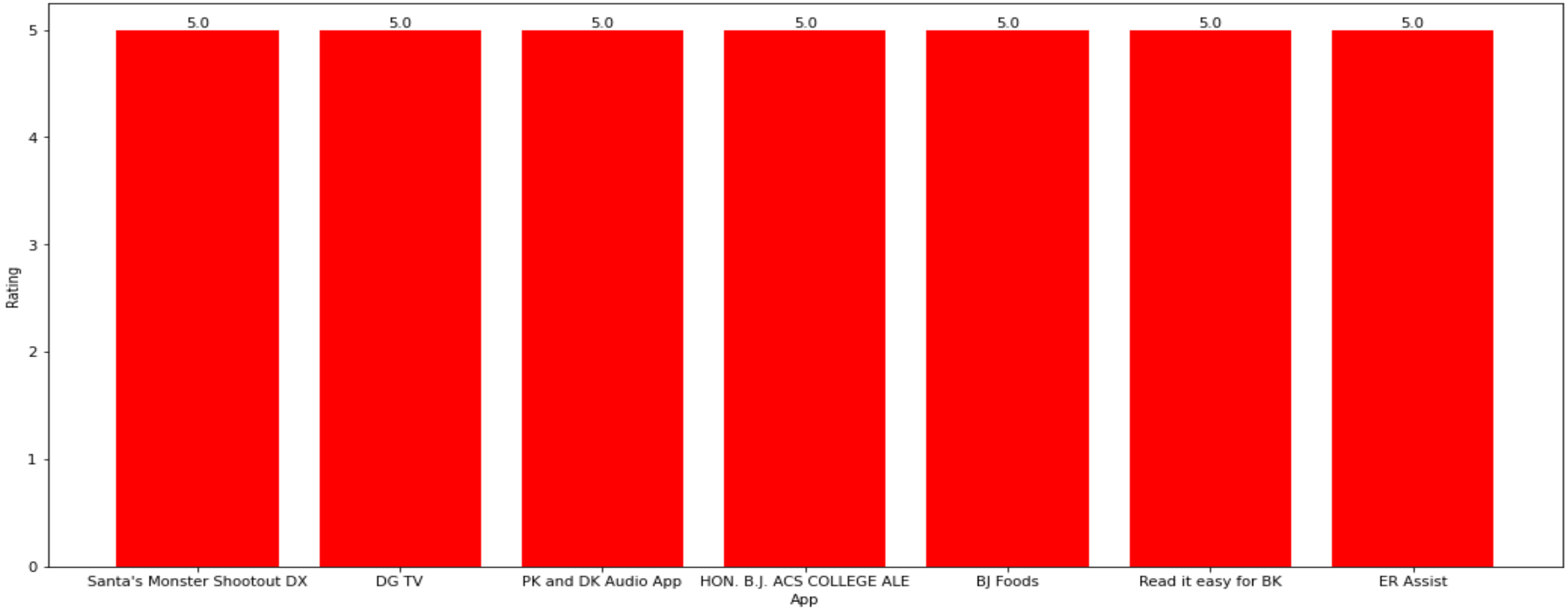
The largest number of installs?



The most popular category



The highest rating



Thanks