

APPLE

App Store

EXPLORATORY
DATA ANALYSIS

OVERVIEW



IDENTIFY STAKEHOLDERS



WHY EDA



FINDING THE INSIGHTS



FINAL RECOMMENDATIONS



SITUATION

Apple App Store Overview:

- Launched in 2008, the Apple App Store has revolutionized the distribution and consumption of mobile applications.
- Hosting millions of apps across categories like games, productivity, and education, it serves as a crucial platform for developers and users worldwide.
- The App Store has significantly contributed to the growth of the mobile app ecosystem, generating billions in revenue and providing countless opportunities for developers.



TASK

Project Objective:

- To conduct an Exploratory Data Analysis (EDA) on Apple App Store data using SQL.
- The goal is to uncover trends, patterns, and key insights within the app marketplace.
- Understanding these factors will aid developers and marketers in making data-driven decisions to enhance app success.



ACTION

Using SQL for EDA:

- Extracting and loading the Apple App Store dataset into a SQL database.
- Cleaning and preprocessing the data to handle missing values, duplicates, and inconsistencies.
- Executing various SQL queries to perform descriptive and exploratory analysis on the dataset.



#1

IDENTIFY THE STAKEHOLDER

Stakeholder is the personal group who has an interest in the outcome of the analysis

So in our case our stakeholder is an aspiring app developer who needs data driven insights to decide what type of data to build and so they are seeking answers to questions like

1. What app categories are most popular?
2. What price should I set?
3. How can I maximize user ratings?



#2

EXPLORATORY DATA ANALYSIS

Exploratory Data Analysis (EDA) is the process of analyzing datasets to summarize their main characteristics, often with visual methods. It helps in understanding the data better and prepares it for further analysis.

1. Get an overview if the apps ratings.
2. Determine if Paid apps have higher ratings than Free.
3. Does App that support more language have higher ratings?
4. Determine Genre with low rating.
5. Check correlation with app description and user ratings.
6. Check Top rated apps for each category.



#3

FINDING THE INSIGHTS



UNIQUE APPLICATIONS ON APPLE APP STORE



Apple_AppStore.eda

```
1 -- CHECKING THE NUMBER OF UNIQUE APPS IN BOTH TABLES  
2 SELECT COUNT(DISTINCT id) AS UniqueAppIDs  
3 FROM AppleStore;  
  
4 SELECT COUNT(DISTINCT id) AS UniqueAppIDs  
5 FROM appleStore_description_combined;
```

⋮ UniqueAppIDs

7197

⋮ UniqueAppIDs

28350

FINDING EMPTY DATA IN DATASET



Apple_AppStore.eda

```
1 -- CHECK FOR ANY MISSING VALUES IN THE FEILDS  
2 SELECT  
3     COUNT(*) AS MISSINGVALUES  
4 FROM  
5     AppleStore  
6 WHERE  
7     track_name IS NULL  
8     OR user_rating IS NULL  
9     OR prime_genre IS NULL;
```

MISSINGVALUES

0

APPLICATIONS IN EACH CATEGORY



Apple_AppStore.eda

```
1 -- FIND OUT THE NUMBER OF APP PER GENRE
2
3     SELECT
4         prime_genre,
5             COUNT(*) AS NumApps
6     FROM
7         AppleStore
8     GROUP BY
9         prime_genre
10        ORDER BY
11            NumApps DESC;
```

prime_genre	NumApps
Games	3862
Entertainment	535
Education	453
Photo & Video	349
Utilities	248
Health & Fitness	180
Productivity	178
Social Networking	167
Lifestyle	144
Music	138
Shopping	122
Sports	114
Book	112
Finance	104
Travel	81

RATING STATISTICS



Apple_AppStore.eda

```
1 -- Get an overview of the apps ratings
2 SELECT
3     min(user_rating) AS MinRating,
4     max(user_rating) AS MaxRating,
5     avg(user_rating) AS AvgRating
6 FROM
7     AppleStore;
```

	MinRating	MaxRating	AvgRating
	0	5	3.526955675976101

RATING STATISTICS FOR FREE AND PAID APPLICATIONS



Apple_AppStore.eda

```
1 -- DETERMINE WHETHER PAID APPS HAVE HIGHER RATING THAN FREE APPS
2 SELECT CASE
3     WHEN price>0 THEN 'PAID'
4     ELSE 'FREE'
5 END AS App_Type,
6 round(AVG(user_rating),2) AS Avg_Rating
7 FROM
8     AppleStore
9 GROUP BY
10    App_Type;
```

App_Type	Avg_Rating
FREE	3.38
PAID	3.72

RATING STATISTICS FOR NUMBER OF LANGUAGE SUPPORT



Apple_AppStore.eda

```
1 -- CHECK IF APPS WITH MORE SUPPORTED LANGUAGES HAVE HIGHER RAINGS
2 SELECT CASE
3     WHEN lang_num <10 THEN '<10 Languages'
4         WHEN lang_num BETWEEN 10 AND 30 THEN '10-30 Languages'
5         ELSE '>30 Languages'
6     END AS language_bucket,
7     round(AVG(user_rating),2) AS Avg_Rating
8 FROM AppleStore
9 GROUP BY language_bucket
10 ORDER BY Avg_Rating DESC;
```

language_bucket	Avg_Rating
10-30 Languages	4.13
>30 Languages	3.78
<10 Languages	3.37

RATING STATISTICS GENRE WISE



Apple_AppStore.eda

```
1 -- CHECK GENRES WITH LOW RATINGS
2 SELECT prime_genre,
3     round(avg(user_rating),2) AS Avg_Rating
4 FROM AppleStore
5 GROUP BY prime_genre
6 ORDER BY Avg_Rating ASC
7 LIMIT 10;
```

prime_genre	Avg_Rating
Catalogs	2.1
Finance	2.43
Book	2.48
Navigation	2.68
Lifestyle	2.81
News	2.98
Sports	2.98
Social Networking	2.99
Food & Drink	3.18
Entertainment	3.25

CORRELATION BETWEEN DISCRIPTION LENGTH AND USER RATING



Apple_AppStore.eda

```
1 -- CHECK IF THERE IS CORRELATION BETWEEN THE LENGTHS OF THE TOP APP
2 DISCRIPTION AND THE USER RAITNG
3 SELECT CASE
4     WHEN length(b. app_desc) < 500 THEN 'Short'
5     WHEN length(b. app_desc) BETWEEN 500 and 1000 THEN 'Medium'
6     ELSE 'Long'
7 END AS description_lenght_bucket,
8 round(avg(a. user_rating),2) AS Average_Rating
9 FROM AppleStore AS a
10 JOIN appleStore_description_combined AS b
11 ON a.id=b.id
12 GROUP BY description_lenght_bucket
13 ORDER BY average_rating DESC;
```

description_lenght_bucket	Average_Rating
Long	3.87
Short	3.54
Medium	3.5

TOP RATED APPLICATION OF APP STORE IN EACH GENRE



Apple_AppStore.eda

```
1 -- CHECK THE TOP-RATED APPS FOR EACH GENRE
2 SELECT prime_genre,
3       track_name,
4       user_rating
5 FROM (
6   SELECT
7     prime_genre,
8     track_name,
9     user_rating,
10    RANK() OVER(PARTITION BY prime_genre ORDER BY user_rating DESC, rating_count_tot DESC) AS rank
11  FROM AppleStore
12 ) AS a
13 WHERE a.rank=1
14 LIMIT 5;
```

prime_genre	track_name	user_rating
Book	Color Therapy Adult Coloring Book for Adults	5
Business	TurboScan™ Pro - document & receipt scanner. scan multiple pages and photos to PDF	5
Catalogs	CPlus for Craigslist app - mobile classifieds	5
Education	Elevate - Brain Training and Games	5
Entertainment	Bruh-Button	5

RESULT

Expected Outcomes:

- Gaining a comprehensive understanding of the app landscape in the Apple App Store.
- Identifying popular app categories, factors influencing app ratings, and pricing trends.
- Providing actionable insights and recommendations for app developers and marketers to improve app performance and strategic planning.



#4 FINAL RECOMMENDATIONS

1. Paid apps have better ratings.
2. Apps supporting between 10 and 30 languages have better ratings.
3. Finance and book apps have low ratings.
4. Apps with a longer description have better ratings.
5. A new app should aim for an average rating above 3.5.
6. Games and entertainment have high competition.

