# EdTech App Market Analysis

Introduction:

Descriptive Statistics

This report presents the analysis of an EdTech apps dataset, focusing on descriptive statistics to understand the distribution and central tendencies of key variables such as installs, scores, ratings, and reviews.

| Metric	Installs	Score	Ratings	Reviews
Mean	5,079,426	3.97	315,978	10,044
Median	100,000	4.16	42,862	89
Mode	1,000,000	3.12	248,487	89
Std Dev	17,240,418	0.65	1,401,577	66,208

App Installs:

- **Distribution:** The distribution of app installs is skewed to the right, with a long tail of apps having a very high number of installs. This suggests that a small number of apps are extremely popular, while the majority of apps have relatively few installs.

- **Range:** The range of installs is vast, spanning from a few thousand to tens of millions.
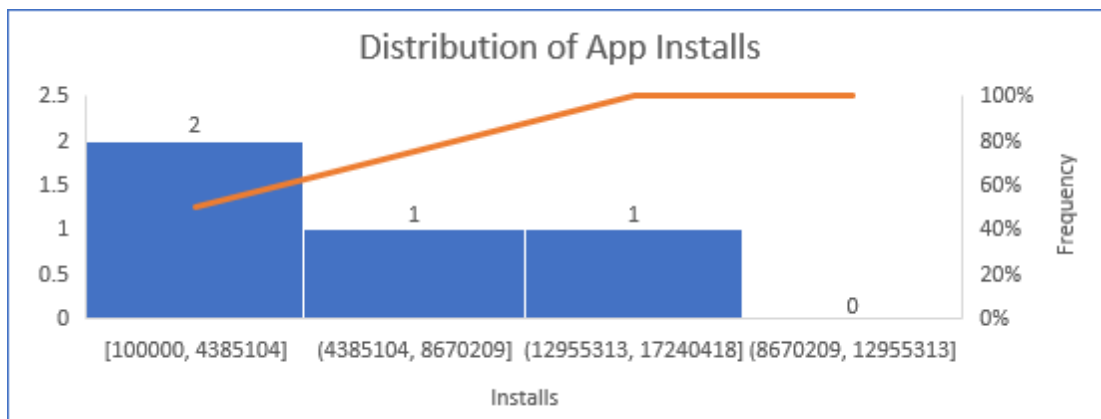
App Scores:

- **Distribution:** The distribution of app scores is relatively symmetric, with a slight skew to the left. This indicates that most apps have scores around the average, with a few outliers with very high or very low scores.

- **Range:** The range of app scores is relatively narrow, suggesting that most apps receive similar ratings from users.

Distribution of App Scores
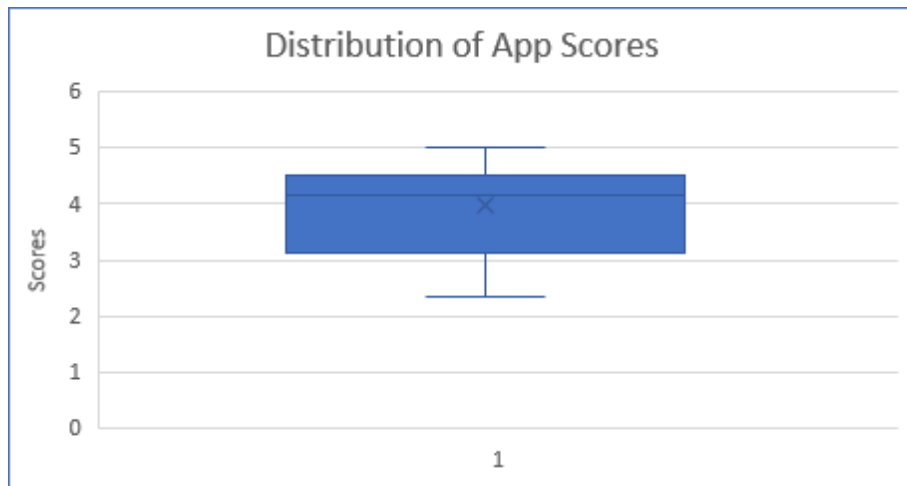
Average Score by Content Rating:

- **Teen content:** Apps rated for "Teen" have the highest average score (4.17), followed by "Everyone" (3.96). This suggests that apps with a broader audience tend to receive slightly lower ratings.

- **Grand Total:** The grand total average score is 3.97, indicating that overall, apps tend to receive positive ratings.

Average Score by Content Rating	
Row Labels	**Average of score**
Everyone	3.96
Teen	4.17
Grand Total	**3.97**

SQL Insights:

Query 1: Top 10 Apps by Installs

```
SELECT TOP 10
    App_Name,
    Installs
FROM
    EdTechApps
ORDER BY
    Installs DESC;
```

Result

App Name	Installs
Duolingo: language lessons	100000000
BYJU'S – The Learning App	100000000
Unacademy Learner App	50000000

IIT JEE, NEET, NCERT Solutions	50000000
Quizlet: Languages & Vocab	10000000
Edmodo	10000000
Cambly - English Teacher	10000000
Noon Academy – Student Learning App	10000000
Teachmint - The Classroom App	10000000
Sololearn: Learn to Code	10000000

Insight: Language learning apps (e.g., Duolingo) and exam preparation apps (e.g., BYJU'S) dominate the top installs, indicating high demand for these categories in the EdTech market.

Query 2: Average Score of Apps with Ads vs. Without Ads

```
SELECT
    containsAds,
    AVG(Score) AS Average_Score
FROM
    EdTechApps
GROUP BY
    containsAds;
```

Result

containsAds	Average_Score
FALSE	3.9
TRUE	4.1

Insight: Apps **with ads** have a slightly higher average score (4.1) than apps without ads (3.9). This challenges the assumption that ads negatively impact user satisfaction, suggesting other factors (e.g., app quality, utility) may offset ad-related drawbacks.

Query 3: Most Popular Android Version

```
SELECT
    Android_Version,
    COUNT(*) AS Number_of_Apps
FROM
    EdTechApps
GROUP BY
    Android_Version
ORDER BY
    Number_of_Apps DESC;
```

Android Vesrion	No . Of apps
5	40
6	9

4.1	8
4	7
4.4	5
7	4
5.1	4
4.2	2
2.3.3	1
8	1

Insight: Android **version 5** is the most widely targeted (40 apps), followed by newer versions like 6 and 7. Developers prioritize backward compatibility, as older versions (e.g., 4.1, 4.4) still hold relevance.

```
SELECT
    Top 6 developer,
    COUNT(*) AS Number_of_Apps
FROM
    EdTechApps
GROUP BY
    developer
ORDER BY
    Number_of_Apps DESC;
```

Developer	Number of Apps
EduRev: Learning, Mock Test & Exam Preparation App	5
BYJU'S	2
TeamLease EdTech Ltd.	2
The Teaching Company	1
Thrive EdTech	1
Udemy	1

Insight: **EduRev** leads with 5 apps, all focused on exam preparation and mock tests. This highlights a strategic focus on niche educational needs (e.g., competitive exams) to capture market share.

Conclusion

The analysis reveals:

1. **Market Dominance**: Language learning and exam prep apps are the most popular, with Duolingo and BYJU'S leading in installs.

2. **Ad Strategy**: Apps with ads perform slightly better in user ratings, suggesting ads can coexist with quality experiences.

3. **Android Prioritization**: Developers should continue supporting Android 5 for compatibility while gradually adopting newer versions.

4. **Niche Focus**: EduRev's success underscores the value of targeting specific educational needs (e.g., exam prep).

Recommendations:

- Developers should invest in high-demand categories (language learning, exam prep).

- Ads can be strategically used without harming ratings if balanced with app utility.

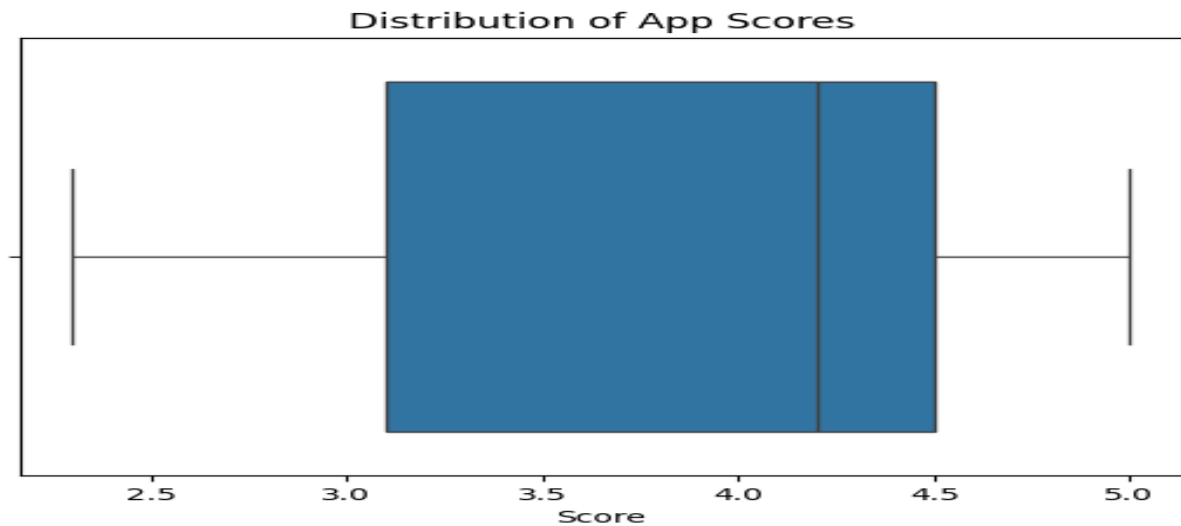- Prioritize Android 5 compatibility while testing newer versions for future-proofing.

Python Analysis

Exploratory Data Analysis (EDA)

Installs Distribution : Most apps have low to medium installs, with a few outliers in the high range.
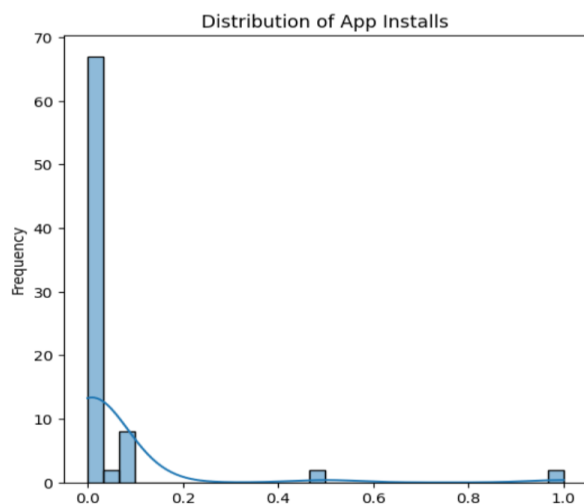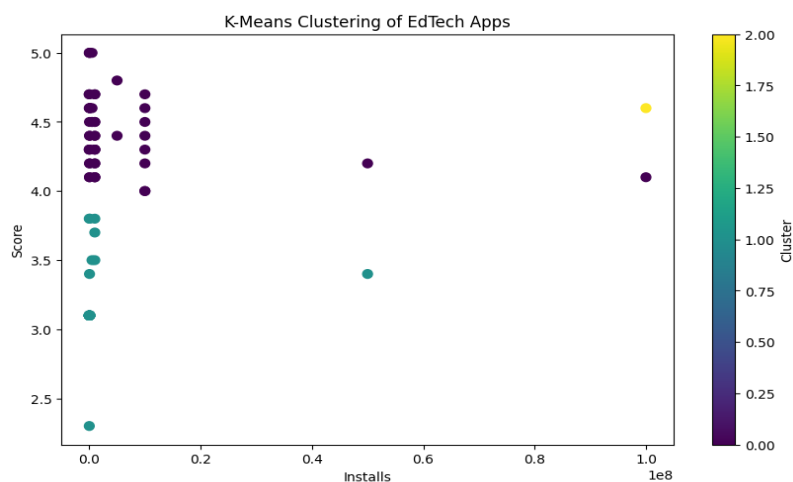

Scores Distribution: App scores are relatively balanced, with most scores between 3.5 and 4.5.

Distribution of App Scores

K-Means Clustering

- **Model**: K-Means Clustering.
- **Features**: Installs, reviews, and app scores
- **Number of Clusters** : 3.

K-Means Clustering of EdTech Apps

Key Findings

- Cluster 0 : Apps with low installs and low scores (e.g., niche or new apps).
- Cluster 1 : Apps with medium installs and medium scores (e.g., moderately popular apps).
- Cluster 2: Apps with high installs and high scores (e.g., top-performing apps like Duolingo and BYJU'S).

Insights

Clustering helps identify distinct groups of apps based on their performance metrics.High-performing apps (Cluster 2) dominate the market, while low-performing apps (Cluster 0) struggle to gain traction.

Recommendations

- Developers should focus on improving app quality and marketing strategies to move apps from Cluster 0 to Cluster 1 or 2.
- Analyze top-performing apps (Cluster 2) to identify best practices for success.

Power BI Dashboard

Dashboard live link – (https://shorturl.at/V3HUM)

Dashboard Overview :

Purpose : Visualize key insights from the EdTech apps dataset.

Components :

- Key metrics (total apps, average score, total installs).

- Top 10 apps by installs.

- Average score by content rating.

- Clusters of apps based on installs and scores.

- Trend of app releases over time.

Insights

Top Apps : Duolingo and BYJU'S dominate in installs.

Content Rating : Teen-rated apps have the highest average score.

Clusters : Apps are grouped into low, medium, and high performers based on installs and scores.

Trends : App releases have increased over time, with a peak in recent years.

Conclusion

1. **Comprehensive Analysis**:

   o This project provided a **holistic analysis** of the EdTech app market, covering key metrics such as installs, scores, ratings, and reviews. Through **descriptive statistics**, **SQL queries**, **Python clustering**, and **Power BI visualizations**, we uncovered actionable insights into app performance and user preferences.

2. **Key Insights**:

   o The analysis revealed that **language learning and exam prep apps** (e.g., Duolingo, BYJU'S) dominate the market in terms of installs. Apps rated

for **"Teen" audiences** tend to have higher average scores, while **ads do not significantly harm user satisfaction**. Clustering further highlighted distinct groups of apps based on their performance, helping identify **low, medium, and high performers**.

3. **Actionable Recommendations**:

   o Developers should focus on **improving app quality** and **targeting high-demand categories** (e.g., language learning, exam prep). Additionally, **optimizing for popular Android versions** (e.g., Android 5) and leveraging **data-driven strategies** can enhance app success. The insights from clustering can guide efforts to move low-performing apps into higher-performing categories.

4. **End-to-End Data Analytics Workflow**:

   o This project demonstrated a **complete data analytics workflow**, from **data cleaning** and **exploratory analysis** in Excel, to **advanced querying** in SQL Server, **AI/ML implementation** in Python, and **interactive dashboard creation** in Power BI. The integration of these tools showcases the ability to transform raw data into **actionable business insights**.