**DATA SCIENCE MINOR PROJECT REPORT**

(Project Semester August-December 2020)

***(GOOGLE PLAY STORE)***

Submitted by

Ajay Kumar

Registration No. 11801375

Programme and section: B.Tech (CSE) K18CJ

Course Code. INT217

Under the Guidance of

**Vasudha**

**Discipline of CSE/IT**

**Lovely School of Computer Science and Engineering**

**Lovely Professional University, Phagwara**

**CERTIFICATE**

This is to certify that Ajay Kumar bearing Registration no. 11801375 has completed INT217 project titled, **“GOOGLE PLAY STORE”** under my guidance and supervision. To the best of my knowledge, the present work is the result of his original development, effort and study.

**Signature and Name of the Supervisor**

**Designation of the Supervisor**

**School of Computer Science and Engineering**

Lovely Professional University

Phagwara, Punjab.

Date: /11/2020

**DECLARATION**

I, Ajay Kumar student of B.Tech (CSE) under CSE Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: /11/2020 Signature

Registration No. 11801375 Ajay Kumar

**ACKNOWLEDGEMENT**

I would like to express my thanks to the people who have helped me most throughout my project. I am grateful to my teacher M/s Vasudha for nonstop support for the project.

A special thank of mine goes to my colleague who helped me out in completing the project, where they all exchanged their own interesting ideas, thoughts and made this possible to complete my project with all accurate information. I wish to thank my parents for their personal support or attention who inspired me to go my own way.

At last but not the least I want to thank my friends who treasured me for my hard work and encouraged me and finally to God who made all the things possible for me till the end.

**TABLE OF CONTENT**

1. Introduction

2. Scope of the Analysis

3. Existing System

1. Drawbacks or limitations of existing system

4. Source of dataset

5. ETL process

6. Analysis on dataset (for each analysis)

1. Introduction
2. General Description
3. Specific Requirements, functions and formulas
4. Analysis results
5. Visualization

7. List of Analysis with results

8. Future scope

9. References

10. Bibliography

**INTRODUCTION**

Google Play (previously Android Market) is a digital distribution service operated and developed by Google. It serves as the official app store for the Android operating system, allowing users to browse and download applications developed with the Android software development kit (SDK) and published through Google. Google Play also serves as a digital media store, offering music, magazines, books, movies, and television programs. It previously offered Google hardware devices for purchase until the introduction of a separate online hardware retailer, Google Store, on March 11, 2015.

Applications are available through Google Play either free of charge or at a cost. They can be downloaded directly on an Android device through the Play Store mobile app or by deploying the application to a device from the Google Play website. Applications exploiting hardware capabilities of a device can be targeted to users of devices with specific hardware components, such as a motion sensor (for motion-dependent games) or a front-facing camera (for online video calling). The Google Play store had over 82 billion app downloads in 2016 and has reached over 3.5 million apps published in 2017.It has been the subject of multiple issues concerning security, in which malicious software has been approved and uploaded to the store and downloaded by users, with varying degrees of severity

Play Store is Google's official pre-installed app store on Android-certified devices. It provides access to content on the Google Play Store, including apps, books, magazines, music, movies, and television programs.

Play Store filters the list of apps to those compatible with the user's device. Developers can target specific hardware components (such as compass), software components (such as widget), and Android versions (such as 7.0 Nougat). Carriers can also ban certain apps from being installed on users' devices, for example tethering applications.

There is no requirement that Android applications must be acquired using the Play Store. Users may download Android applications from a developer's website or through a third-party app store alternative. Play Store applications are self-contained Android Package files (APK), similar to .exe files to install programs on Microsoft Windows computers. On Android devices, an "Unknown sources" feature in Settings allows users to bypass the Play Store and install APKs from other sources. Depending on developer preferences, some apps can be installed to a phone's external storage card.

Android users have complained that the Google Play store access cannot be blocked and there is constant data exchange with the google cloud. Also valuable CPU resources are used, slowing down the Android syste

**SCOPE OF ANALYSIS**

The system provides different types of services based on the different types of application in the play store. The analysis provides we count the downloaded application in each category wise number of software that costumer had downloaded and which software has most downloaded.

The feature that are available in play store data management are:

* It provides overall download of every application.
* It provides the current update every month.
* It provides overall profit per year or the shopkeeper can check the profit on different book per month.
* It provides quantity of different application with their sum of install.
* It provides sum of different types of category of application.
* It provides the information top downloaded application
* It provides the top review in each category.
* It provides the top rating wise in each category.
* It provides the number of applications downloaded.
* It provides top downloaded category wise application.

By this analysis costumer know that Health & Fitness, Travel & Local and Education were the most common categories, accounting for about 15%, 14% and 13%, respectively, of the total number of apps in our dataset. Food & Drink was the least prevalent category with only 87 observations (less than 1% of the total number of apps in our dataset). Four categories of applications, including Health & Fitness, Travel & Local, Education and Finance, accounted for more than 50% of the apps in our dataset.

.

**My System**

The data looks like this:

Table

Description automatically generated

In the data the columns like **apps, category, rating, reviews, size, installs, type, price, content rating, genres**

**Last update, current version, android version**

**apps:** In this column the data is filled according to total number of installs.

**category:** In this column the data filled category wise.

**rating:** In this column the data filed with increasing to decreasing wise.

**reviews:** In the given data set filled with reviews.

**size:** In this column the number of software ordered by each size.

**install:** In this column total number of software downloaded by each category.

**types:** In this column date filled with given software free or paid.

**price:** In this column different types of price are given.

**Content rating:** In this column the given data set told about who can access the data like everyone10+ or teen.

**genres:** In this column the told about given dataset is video player or photographics.

**Last Update:** This column contains the info of last updated version of an app**.**

**Current Version:** This Column contains the info of current version of an app.

**Android version:** In this column we know which type of android version required for given software.

**Sales:** In this we use formula for counting sales by multiplying selling price by total number of books.

In existing System we can find out

* Total number of software and download
* Sum of software in different category
* Total number of software according to rating
* Each category wise software and reviews
* Total number of software free or paid
* Total number of downloaded applications
* Number of applications according to android version

**Drawbacks or limitations of existing system:**

While looking for the resources to complete this analysis we found many other analysis, but majority of them were wrong since they counted an item twice or thrice which lead to a huge increment on their values and they had also performed their analysis using python or R.

We perform this analysis in Microsoft excel, therefore, it will accessible to everyone because majority of people know how to use excel and can understand the analysis.

. We can’t enter large amount of data in excel sheet manually so we need some software for managing the data sheet. In given data analysis can be done only once after that if data set has changed then for updated data analysis is required again. For managing larger data we use power pivot and one of the limitation of power pivot is it does not provide functionality as compared to excel sheet

**Source of dataset**

The source of data is online site <https://www.kaggle.com/datasets>

Kaggle is an online community of data scientists and machine learners, owned by google inc. Kaggle allows users to find and publish data set, explore and build models in a wed-based data-science environment, work with other data scientists and machine learning engineers

**OBJECTIVE:**

1. CLEASING

Before Cleasing:

Table, Excel

Description automatically generated

After Cleasing:

Table

Description automatically generated

1. ANALYSIS

Graphical user interface, application

Description automatically generatedGraphical user interface, text, application, table

Description automatically generatedGraphical user interface, text

Description automatically generatedA picture containing graphical user interface, table

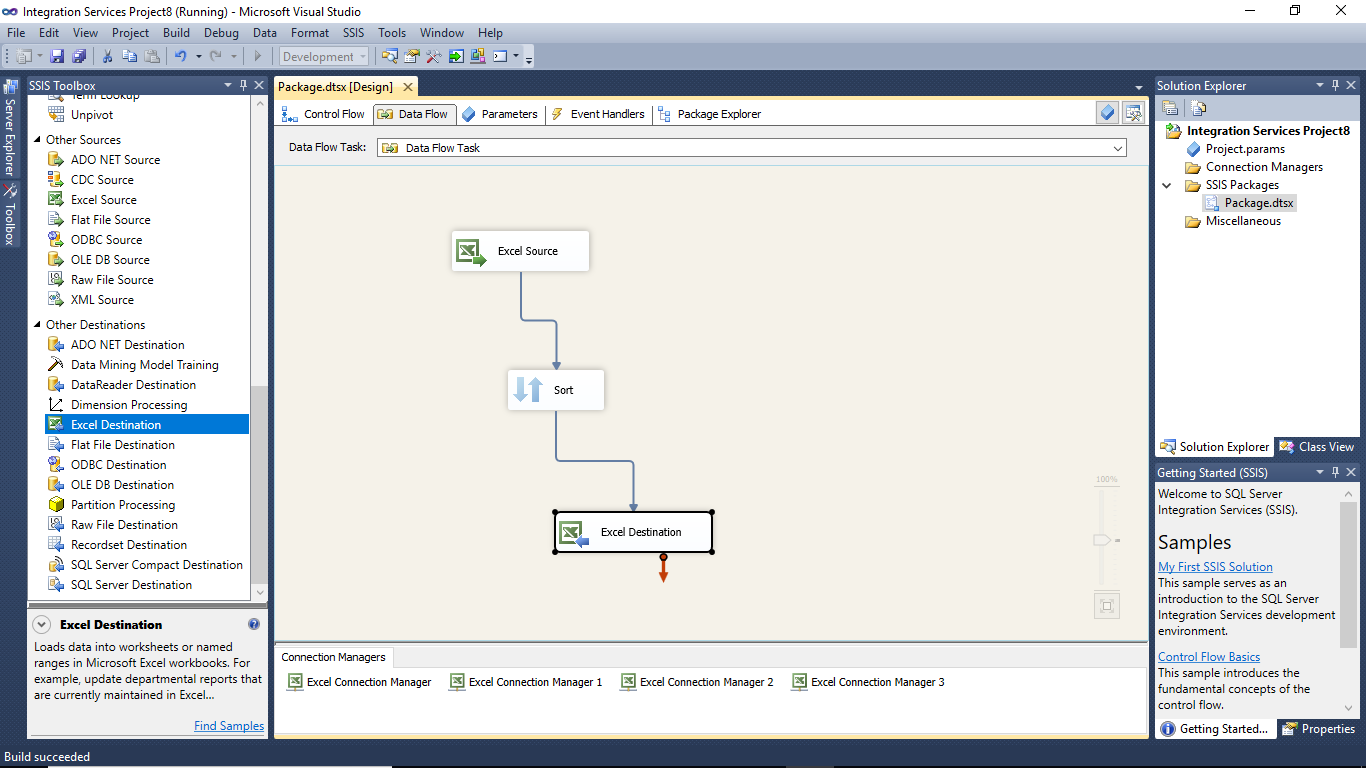
Description automatically generatedTable

Description automatically generated

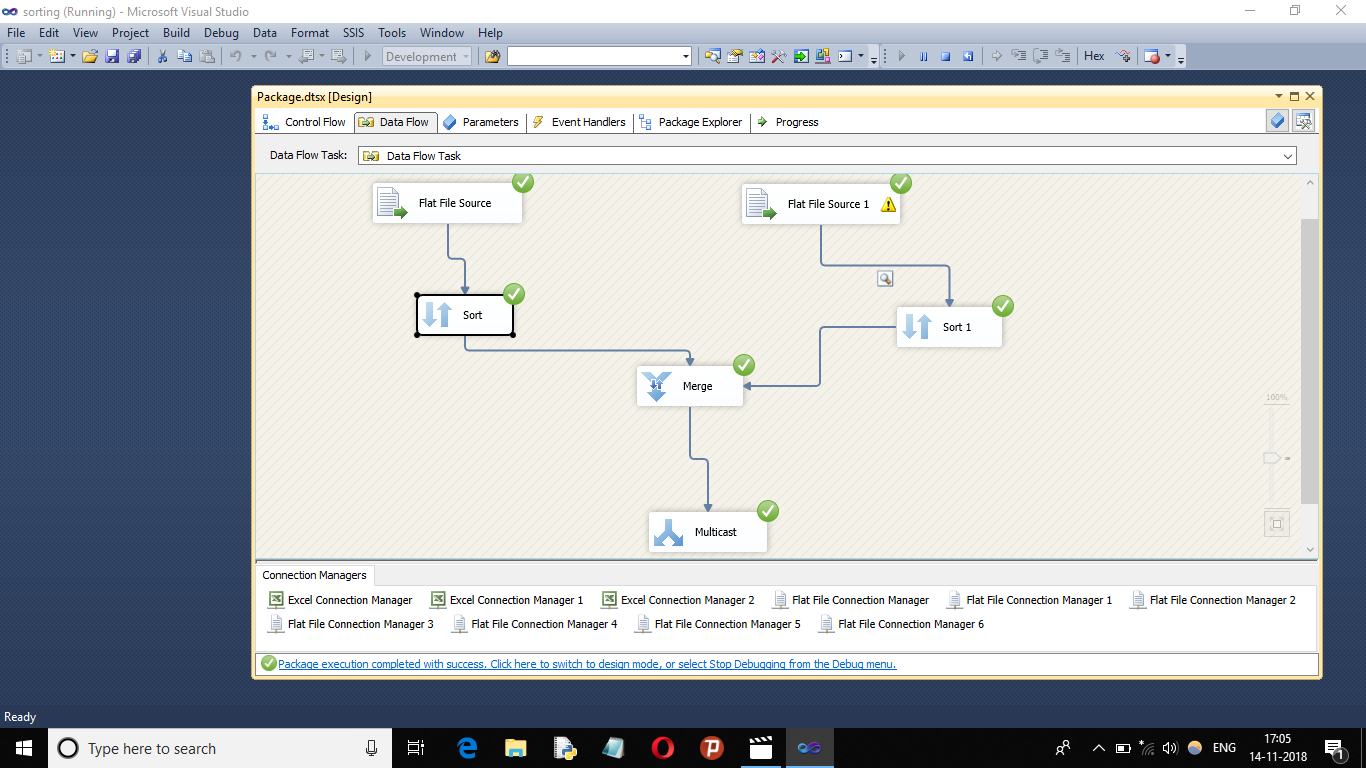
|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

**ETL process**

**Extraction Of Data:**



**Sorting Of Data:-**

****

**Analysis on dataset (for each analysis)**

**Introduction:**

Data analysis is a process of inspecting, [cleansing](https://en.wikipedia.org/wiki/Data_cleansing), [transforming](https://en.wikipedia.org/wiki/Data_transformation), and [modeling](https://en.wikipedia.org/wiki/Data_modeling) [data](https://en.wikipedia.org/wiki/Data) with the goal of discovering useful information, informing conclusions, and supporting decision-making. Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, while being used in different technology, science, and social science domains.

The process of data analysis:

* Data requirements
* Data collection
* Data processing
* Data cleaning
* Modeling and algorithms

**Specific Requirements, functions and formulas:**

We use different formulas for data cleaning, data processing and modelling and algorithms.

Main function of excel that I have used are

* Power pivot
* Pivot table
* Different arithmetic formulas
* Conditional formatting
* Create graph using power pivot
* Create relation between different table
* Function analyser (Slicer)
* Different types of graph
* Different functionality of graph
* Comparison between different quantities
* Absolute cell reference
* Cell formatting

**ANALYSIS RESULT:**

Analysis result are mention below:

**VISUALIZATION**

1. **Which type of content have more software**

**Introduction**

In this analysis we tried to find out the name of content rating which is most imported the products on basis on quantity.is content more software

Specific requirements, functions and formulae:

This does not require any special functions or formulas. This was done by consolidating the dataset into a pivot table. We used the pivot table and then visualized it in the graphs given below.

* **Android version:**
* **Introduction**

In this analysis we tried to find out the android version and total number of install

Specific requirements, functions and formulae:

This does not require any special functions or formulas. This was done by consolidating the dataset into a pivot table. We used the pivot table and then visualized it in the graphs given below.

**Total number of software in each Genre:**

* **Introduction**

In this analysis we tried to find out the total number of software in each content rating

Specific requirements, functions and formulae.

This does not require any special functions or formulas. This was done by consolidating the dataset into a pivot table. We used the pivot table and then visualized it in the graphs given below

**Total number of free software and paid in each category:**

* **Introduction**

In this analysis we tried to find out the total number of free software and paid software

Specific requirements, functions and formulae.

This does not require any special functions or formulas. This was done by consolidating the dataset into a pivot table. We used the pivot table and then visualized it in the graphs given below

**Number of install software vs size of software:**

* **Introduction**

In this analysis we tried to find out the total number install software according to there rating wise

Specific requirements, functions and formulae.

This does not require any special functions or formulas. This was done by consolidating the dataset into a pivot table. We used the pivot table and then visualized it in the graphs given below

**Reviews:**

* **Introduction**

In this analysis we tried to find out the reviews genre wise

Specific requirements, functions and formulae.

This does not require any special functions or formulas. This was done by consolidating the dataset into a pivot table. We used the pivot table and then visualized it in the graphs given below

**Visualization:**

Data visualization is viewed by many disciplines as a modern equivalent of [visual communication](https://en.wikipedia.org/wiki/Visual_communication). It involves the creation and study of the [visual](https://en.wikipedia.org/wiki/Visual_system) representation of [data](https://en.wikipedia.org/wiki/Data).

To communicate information clearly and efficiently, data visualization uses [statistical graphics](https://en.wikipedia.org/wiki/Statistical_graphics), [plots](https://en.wikipedia.org/wiki/Plot_(graphics)), information graphics and other tools. Numerical data may be encoded using dots, lines, or bars, to visually communicate a quantitative message. Effective visualization helps users analyze and reason about data and evidence. It makes complex data more accessible, understandable and usable. Users may have particular analytical tasks, such as making comparisons or understanding causality, and the design principle of the graphic (i.e., showing comparisons or showing causality) follows the task. Tables are generally used where users will look up a specific measurement, while charts of various types are used to show patterns or relationships in the data for one or more variables.

**Future Scope**

Data analytics is a process through which data is cleaned, analyzed and modelled using tools. This data is then used to derive insights. The insights are then used for technology related decision-making purposes. There are many techniques that data analysts use in different fields of work. In the world of technology, Data analytics is used for making strategies to get the desired technology results. Today, data analytics has become a big career option in India. As a result, big data analytics courses are in huge demand.

Technology have realized the importance of utilizing big data analytics to maximize their profits. They know that it is vital for their growth and for the future health of their technology. Today, major technology decisions are taken by utilizing the insights derived from data related to the organization or industry related data. As competition increases and customers are flooded with choices, it has become important to move faster in the market and that too with accuracy.

**Bibliography**

* <https://nycdatascience.com/blog/student-works/web-scraping/analysis-of-apps-in-the-google-play-store/>
* <https://www.kaggle.com/datasets>
* <https://www.youtube.com/results?search_query=dashboard+on+google+play+store+in+excel>

**OBJECTIVE**

1. **Which category have more software?**
2. **Different type of Android version**
3. **Total number of software in each content rating**
4. **Total number of free software and paid**
5. **Total number of install software**
6. **What is last update of software?**
7. **Overall review**