Dataset ‘’ Video games ‘’ :

Description :

This dataset contains a list of video games with sales greater than 100,000 copies. It was generated by a scrape of vgchartz.com.

Fields include

- Rank - Ranking of overall sales
- Name - The games name
- Platform - Platform of the games release (i.e. PC,PS4, etc.)
- Year - Year of the game's release
- Genre - Genre of the game
- Publisher - Publisher of the game
- NA_Sales - Sales in North America (in millions)
- EU_Sales - Sales in Europe (in millions)
- JP_Sales - Sales in Japan (in millions)
- Other_Sales - Sales in the rest of the world (in millions)
- Global_Sales - Total worldwide sales.

The script to scrape the data is available at https://github.com/GregorUT/vgchartzScrape.

It is based on BeautifulSoup using Python.

There are 16,598 records. 2 records were dropped due to incomplete information.

Dataset ‘’ Google Play Store ‘'

Description :

Each app (row) has values for category, rating, size, and more.

The Play Store apps data has enormous potential to drive app-making businesses to success. Actionable insights can be drawn for developers to work on and capture the Android market!

Fields include

- App
- Category
- Rating
- Reviews
- Size
- Installs
- Type
- Price
- Content rating
- Genres
- Last updated
- Current version
- Android version

Dataset ‘’Car Sales ‘'

Description :

This is the Car sales data set which include information about different cars . This data set is being taken from the Analytixlabs for the purpose of prediction

In this we have to see two things

First we have see which feature has more impact on car sales and carry out result of this

Secondly we have to train the classifier and to predict car sales and check the accuracy of the prediction.

Fields include

- Manufacturer
- Model
- Sales in thousands
- Year resale value
- Vehicle type
- Price in thousands
- Engine size
- Horse power
- Wheel base
- Width
- Length
- Curb weight
- Fuel capacity
- Fuel efficiency
- Latest launch
- Power perfFactor

We used this commands to run spark in Google Colab without install in our computers.