### Capstone Two: Project Ideas & Proposal

# Google Play Store apps and Mobile Apple Store Statistics

Analyze Google and Apple store data independently and try to correlate it.

* Most popular app categories
* See dependencies of content (age group) and downloads
* See the genre proportion (tools/entertainments/books/games etc.)
* Compare statistics for different groups
* Check if I can compare statistics for Android and Apple

GooglePlayStore Apps

<https://www.kaggle.com/lava18/google-play-store-apps?select=googleplaystore.csv>

(10842 records 13 columns)

Mobile Apple Store Statistics

<https://www.kaggle.com/ramamet4/app-store-apple-data-set-10k-apps?select=appleStore_description.csv>

(7174 records 17 columns)

1. LendingClub data available

See if I can create a model to approve loans

(borrower data – anuual income, employment duration, other loans vs loan amount, term, interest, term etc

<https://www.kaggle.com/wordsforthewise/lending-club?select=accepted_2007_to_2018Q4.csv.gz>

<https://www.kaggle.com/wordsforthewise/lending-club?select=rejected_2007_to_2018Q4.csv.gz>

(2 huge datasets 10485576 records 50+ columns – accepted dataset and 10485575 9 columns – rejected dataset)

1. Fake news

[https://www.kaggle.com/mrisdal/fake-news](https://www.kaggle.com/mrisdal/fake-news%20)  (1 dataset)

The dataset contains text and metadata from 244 websites and represents 12,999 posts in total from the 30 days (October-November 2016)

(17949 records 20 columns) <https://www.kaggle.com/mrisdal/fake-news?select=fake.csv>

[https://www.kaggle.com/clmentbisaillon/fake-and-real-news-dataset (2](https://www.kaggle.com/clmentbisaillon/fake-and-real-news-dataset%20(2) datasets)

Fake dataset (17949 records 4 columns) <https://www.kaggle.com/clmentbisaillon/fake-and-real-news-dataset?select=Fake.csv>

True dataset (21418 records 4 columns – title/text/subject/date) <https://www.kaggle.com/clmentbisaillon/fake-and-real-news-dataset?select=True.csv>

1. London Bullion Market Association

(Gold and silver price from 01/01/1968 to today)

[https://www.quandl.com/data/LBMA-London-Bullion-Market-Association (4](https://www.quandl.com/data/LBMA-London-Bullion-Market-Association%20(4) datasets)

Gold Price: London Fixing

1/2/1968 – 1/14/2021 (13407 records 7 columns)

Silver Price: London Fixing

1/2/1968 -1/14/2021 (13417 records 4 columns)

Gold Forward Offered Rates (GOFO)

7/17/1989 – 1/30/2015 (6458 records 16 columns)

London Gold Fixings (1991-1999)

1/2/1991 - 12/30/1999 (2275 records 5 columns)

Here are some questions to consider to help you get started: ● What is the problem you want to solve? ● Who is your client and why do they care about this problem? In other words, what will your client do or decide based on your analysis? ● What data are you using? How will you acquire the data? ● Briefly outline how you’ll solve this problem. Your approach may change later, but this is a good first step to get you thinking about a method and solution. ● What are your deliverables? Typically, this includes code, a paper, or a slide deck. For this capstone, the deliverables include: ○ A GitHub repo containing the work you complete for each step of the project, including: ■ A slide deck ■ A project report

Stackoverflow data

https://www.kaggle.com/stackoverflow/stackoverflow

Neflix Shows

<https://www.kaggle.com/shivamb/netflix-shows>

<https://www.kaggle.com/shivamb/netflix-shows?select=netflix_titles.csv> (6K records)