# Analyzing Google Play Apps to identify most popular app sector
# ETL Project

## Team members

Fenny Patel, Justin Ho, Nikki Ryan, Parisa Charkhgary, Sidney Bowe

## Extract

| File Name | Source | Format |
|---|---|---|
| extended_googleplaystore.csv | kaggle.com<br>Google Play Store Apps Extended<br>Based on the original dataset 'Google Play Store Apps' with more features | CSV |
| extended_googleplaystore_user_reviews.csv | kaggle.com<br>Google Play Store Apps Reviews (+110K Comment)<br>Web scraped data of over 110k reviews from different genres of Apps. | CSV |
| Google-Playstore.csv | kaggle.com<br>Google Play Store Apps<br>Google Play Store App data of 2.3 Million+ applications. | CSV |

## Transformation

| File Name | Data Frame Name | Rejected data | Kept Columns | Renamed Columns |
|---|---|---|---|---|
| extended_googleplaystore.csv | clean_neo<br>(removed 15 columns) | None | App, Genres (categorical), Rating, Reviews, Installs, Type, Price, Content Rating | Genres (categorical)' changed to 'genre<br>All other columns changed to lower case<br>Spaces on columns names replaced with _ |
| extended_googleplaystore_user_reviews.csv | clean_neo_reviews<br>(removed 30 columns) | dropna | app, translated_review, original_sentiment | All columns names changed to lower case. Spaces on columns names replaced with _ |
| Google-Playstore.csv | clean_gauthamp<br>(removed 14 columns) | | app_name, category, rating, rating_count, installs, free, price, content_rating,<br>ad_supported, in_app_purchases | All columns names changed to lower case. Spaces on columns names replaced with _ |

## Load

| Finale DB | Table |
|---|---|
| PostgreSQL using SQLAlchemy | neomatrix |
| PostgreSQL using SQLAlchemy | user_reviews |
| PostgreSQL using SQLAlchemy | gauthamp |

## Analysis

The data has been prepared to review GoogleApps ratings to …