ETL of Netflix Data

Extract

Start

CSV Files for
- Netflix Orginal
- Netflix Revenue
- Netflix Subscribers were uploaded from Kaggle

- Files placed in Resource folder
- Used Panda to read files and converted to dataframes.
- Extracted relevant columns such as Title, Genre and Premiere from .csv file.

Transform

Data was also filtered to show only US and Canada Regions

Date field was transformed to "Q1-20XX" Format.

- Data was cleansed of any NAN values
- Data was transformed to match other tables such as date fields

Load

- Netflix DB is created in Postgres
- Netflix_originals, Netflix _subscribers and Netflix_revenue tables were created using SQL query/script

Used pandas engine to load the data in Postgres DB

Created a join and a view using SQL to look the netflixoriginal data with subsriber information.

end