

Build and Deploy Recommender System using Streamlit and Heroku

-by Qi Sun

Professor: Andy Catlin

Yeshiva University

Spring 2021

Dataset:

- **Source:** MovieLens dataset
 - This dataset is an ensemble of data collected from TMDB and GroupLens.
- **Movie data:** 45,000 movies
 - Consists of movies released on or before July 2017
 - Features include posters, backdrops, budget, revenue, release dates, languages, production countries and companies.
- **Rating data:**
 - Full dataset: 26 million ratings from 270,000 users for all 45,000 movies
 - Ratings are on a scale of 1-5
 - 'ratings_small.csv': the subset of 100,000 ratings from 700 users on 9,000 movies.

Method:

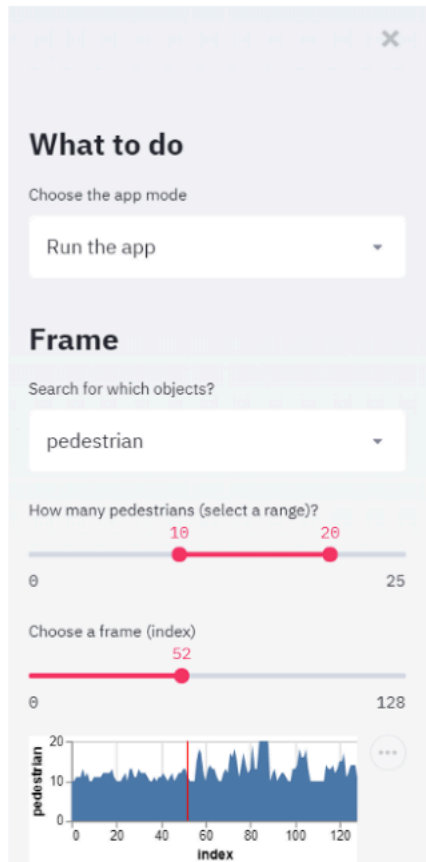
- Content-based filtering
 - Based on the overview of movie
 - Make text lowercase
 - Word tokenize and Remove stop words, punctuation, and words containing numbers
 - Text similarity: cosine similarity
 - CountVectorize
- Collaborative-based filtering
 - Surprise package– SVD
 - Remove the rated movies

Streamlit:

- <https://streamlit.io/>
- Installation: `pip install streamlit`
- Advantages:
 - The fastest way to build and share data apps.
 - Python, No HTML knowledge is needed
 - Open-source and free
 - Create application with only a few lines of code
- Demo: local app

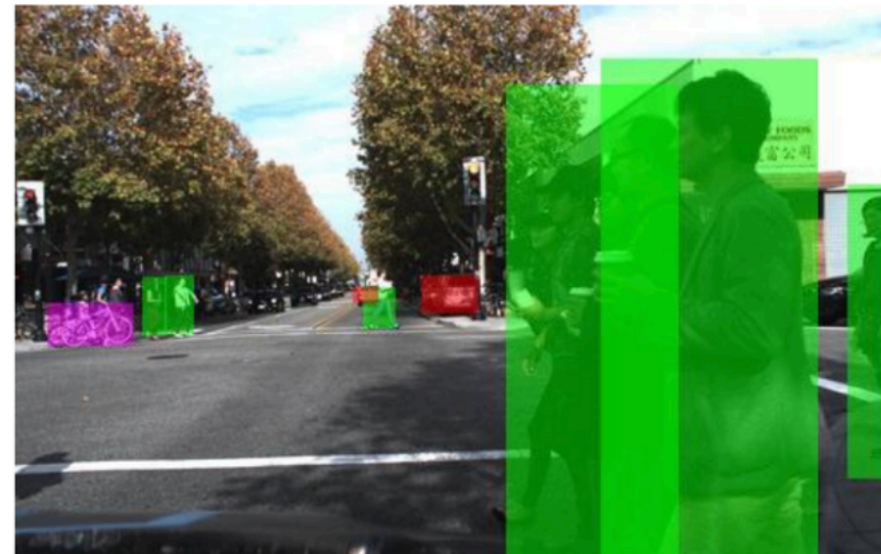
Example:

The Self-driving Car Image Browser



Real-time Computer Vision

YOLO v3 Model (overlap 0.3) (confidence 0.5)



Uber Pickups in New York City

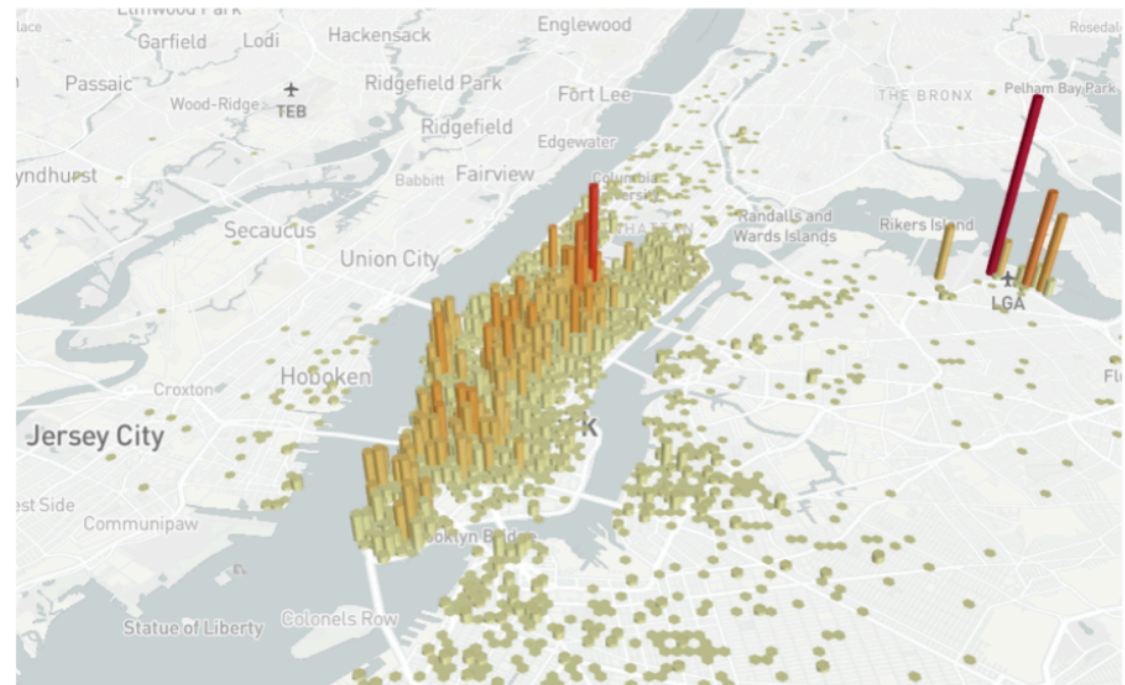
This is a demo of a Streamlit app that shows the Uber pickups geographical distribution in New York City. Use the slider to pick a specific hour and look at how the charts change.

[See source code](#)

Hour to look at



Geo data between 17:00 and 18:00



Heroku:

- Deploy my Streamlit application to Heroku
- Cloud service platform for app deployment and management
- free tier

The screenshot shows the Heroku dashboard for a user named 'susanandtest'. The top navigation bar includes the Heroku logo, a search bar with the text 'Jump to Favorites, Apps, Pipelines, Spaces...', and a user profile icon. Below the navigation bar, the user's profile is shown with a dropdown menu for 'Personal' and a link to 'susanandtest'. There are buttons for 'Open app' and 'More'. The main content area is divided into two sections. The left section, titled 'Add this app to a pipeline', explains that users can create a new pipeline or choose an existing one to add the app to a stage. The right section, titled 'Add this app to a stage in a pipeline to enable additional features', shows two options: connecting the app to a pipeline to enable code promotion, and connecting it to a pipeline connected to GitHub to enable review apps. At the bottom, there is a 'Deployment method' section with three options: 'Heroku Git' (Use Heroku CLI), 'GitHub' (Connected, with a green checkmark), and 'Container Registry' (Use Heroku CLI).

HEROKU

Jump to Favorites, Apps, Pipelines, Spaces...

Personal > susanandtest

GitHub susanqisun/recommender

Open app More

Overview Resources **Deploy** Metrics Activity Access Settings

Add this app to a pipeline

Create a new pipeline or choose an existing one and add this app to a stage in it.

Add this app to a stage in a pipeline to enable additional features

Pipelines let you connect multiple apps together and **promote code** between them. [Learn more.](#)

Pipelines connected to GitHub can enable **review apps**, and create apps for new pull requests. [Learn more.](#)

Choose a pipeline

Deployment method

Heroku Git Use Heroku CLI

GitHub **Connected**

Container Registry Use Heroku CLI

Application:

- App address:
 - <https://susanandtest.herokuapp.com>
- Github address:
 - <https://github.com/susanqisun/recommender/tree/main>

References:

- https://github.com/Jcharis/Streamlit DataScience Apps/blob/master/course_recommendation_sys_app/app.py
- <https://stackoverflow.com/questions/66718228/select-multiple-options-in-checkboxes-in-streamlit>
- <https://gilberttanner.com/blog/turn-your-data-science-script-into-websites-with-streamlit>
- <https://github.com/ChristianFJung/NotebookToWebApp/blob/master/article.md>
- <https://boadziedaniel.medium.com/part-2-the-salary-predictor-a7f4c50a84ca>
- <https://www.analyticsvidhya.com/blog/2020/12/deploying-machine-learning-models-using-streamlit-an-introductory-guide-to-model-deployment/>