Build and Deploy Recommender System using Streamlit and Heroku

-by Qi Sun
Instructor: 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.

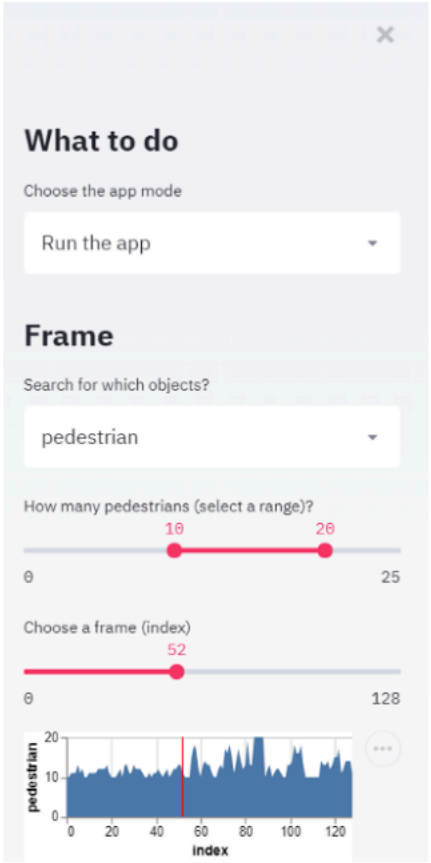
Method:

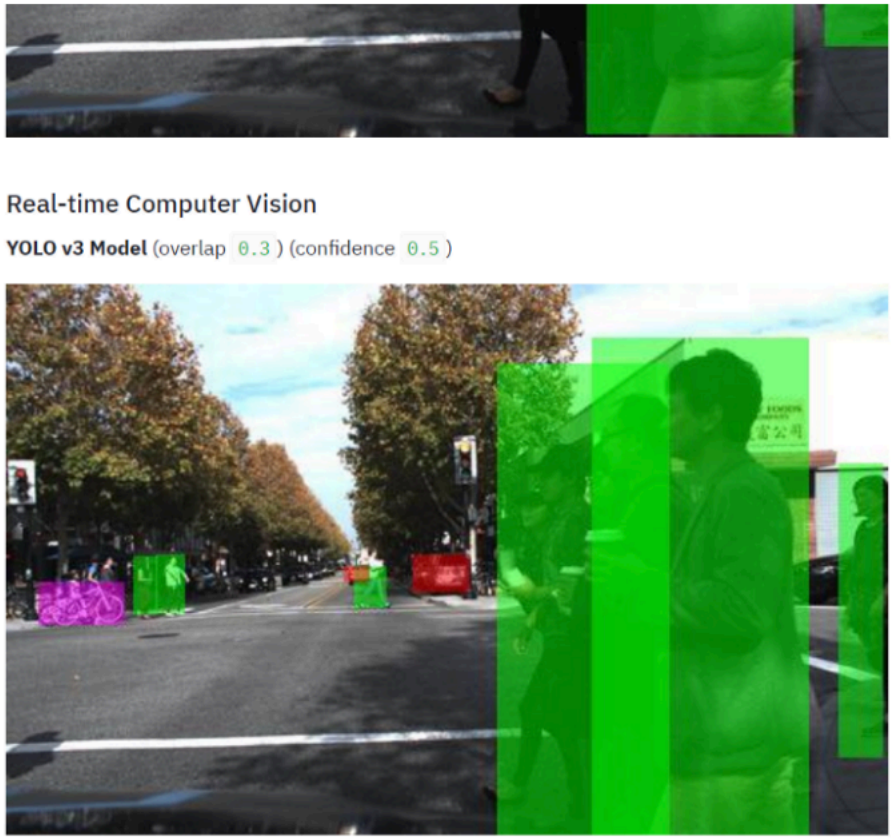
- Content-based filtering
  - Based on the overview of movie
  - Make text lowercase
  - Word tokenize and Remove stop words, punctuation, and words containing numbers
  - Text similarity: cosine similarity
  - CountVectorize
- Collaborative-based filtering
  - Surprise package– SVD
  - Remove the rated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





Reference: https://towardsdatascience.com/quickly-build-and-deploy-an-application-with-streamlit-988ca08c7e83

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.

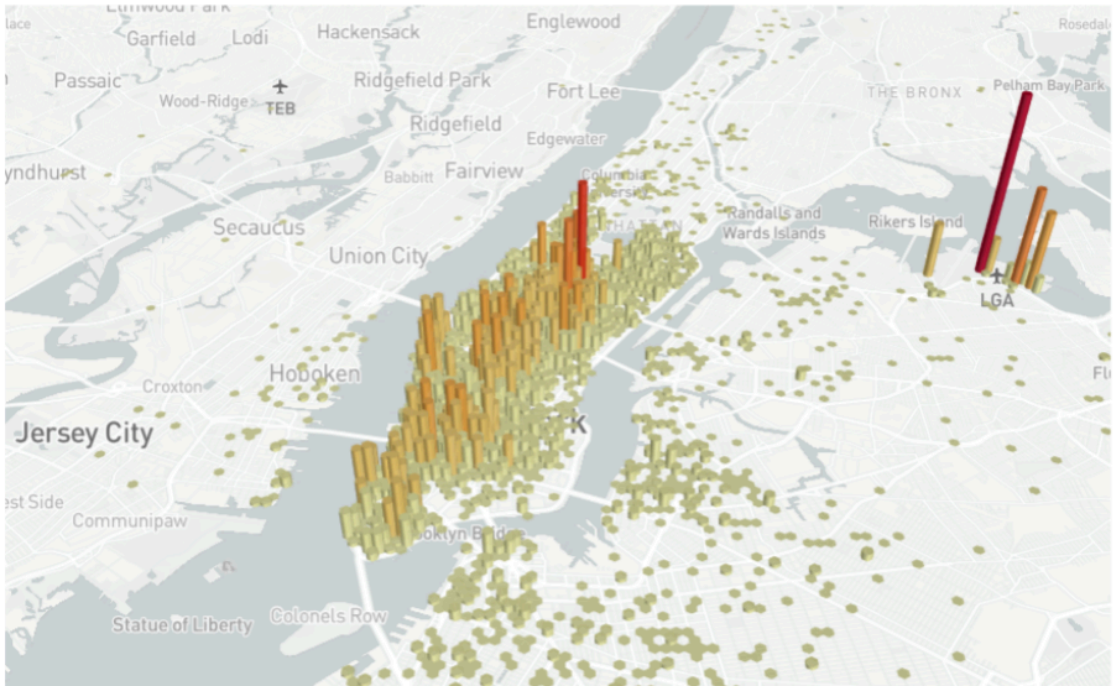
See source code

Hour to look at

17

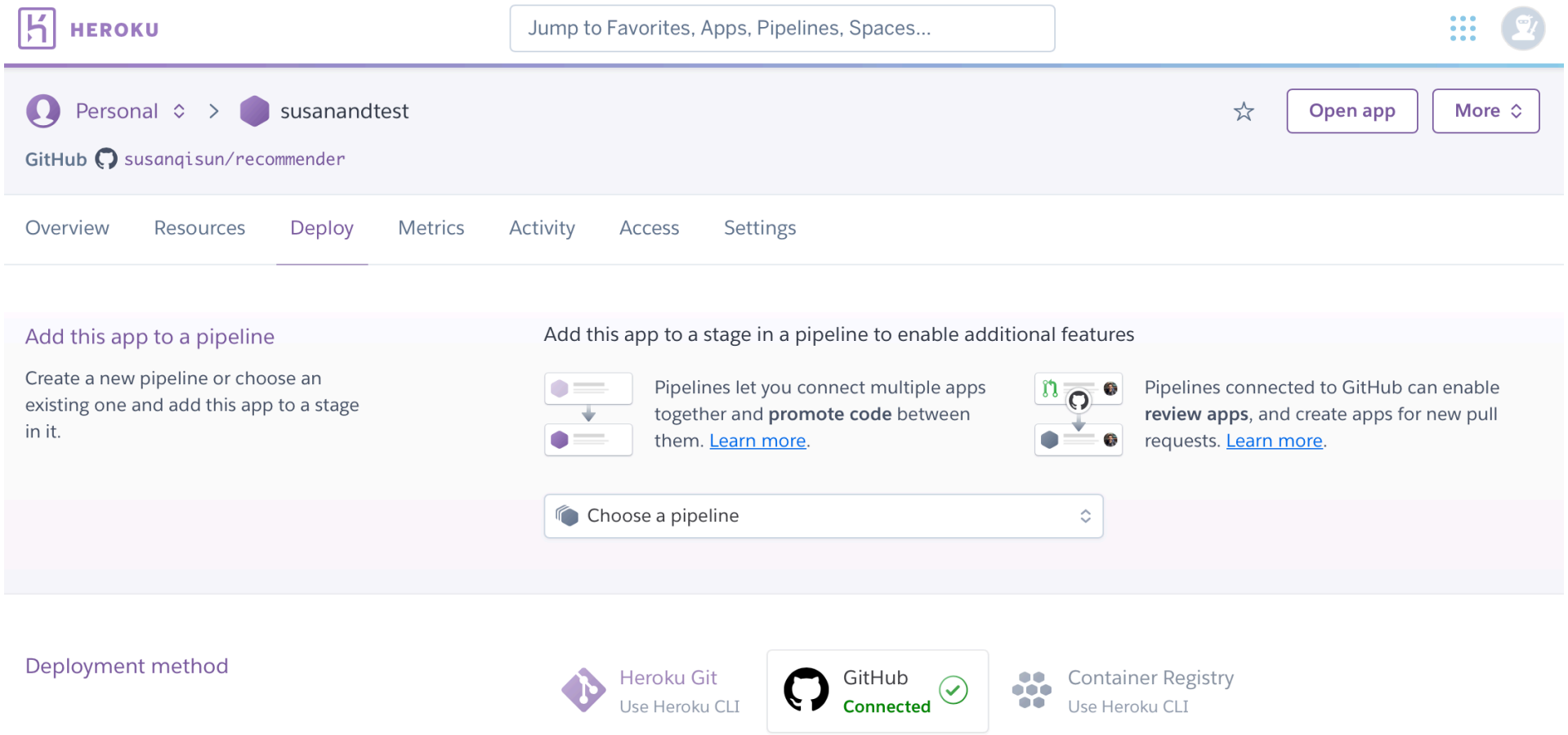
0 23

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



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/