

UNIVERSITY *of* WASHINGTON

Software Design for Data Science

Project Proposals

*DATA 515 Students
University of Washington
February 7, 2023*

W

Proposals & Team Formation

During project proposals:

- Volunteer to present!
- Take notes on what projects sound interesting to you

After project proposals:

- You'll have time in class to talk to each other and form teams around a project proposal
- 3-4 people per team (4 is optimal)
- If a team has 1-2 people and can't find more people or a project, I will help coordinate

Projects

Airbnb Recommender – Top 5 Airbnbs in your Desired City in the US

Airline Crew Pairing/Scheduling

Airline Delay Prediction

Analyzing Academic Sexual Misconduct

Analyzing the Wisconsin Wolf Population

Anomaly Detection in Financial Transactions

BEAT THE CROWD! *

Beers, Breweries, and Better Choices

Car Decision Making Interactive Helper

Charge Me Up!

Climate Change and Wildfire in the US

Projects, Part 2

[COVID19: Patient impact and State policy](#)

[Create an API Overlay for Smogon.com](#)

[Crime Analysis](#) *

[Crime under the Weather](#)

[Customer Personality Analysis](#)

[Data Analysis: Rotten Tomatoes and the Oscars](#)

[Electric Vehicle Charging Heat Map](#)

[Exploring Weather Trends](#)

[Face Extractor Tool](#)

[Food Deal-ivery](#) *

[FRAME - Food Recommendations For All Methodical Eaters](#)

[Global Power Plants and Energy Consumption Analysis](#)

Projects, Part 3

[Impact of Sci-Fi among Fictions over the Years](#)

[Kaggle Genie](#)

[Monte Carlo LVII](#)

[Multilingual OCR for Indic Scripts](#)

[Orphaned Well Mapping Tool](#)

[Parking Spot Finder *](#)

[Pet Palette: Your Guide to Adoptable Paws Heat Map & Breed Comparison](#)

[Popular Video Games](#)

[Recipe Generator](#)

[Relation between crime rates and housing prices in Seattle](#)

[Seattle Weather Prediction](#)

[Sleep Research](#)

Projects, Part 4

Speech To Text Engine

STUDY OF GLOBAL PROGRESS TOWARDS SUSTAINABLE DEVELOPMENT GOALS

Term Rate Conversion for Banks

Tourist Places Recommendation System

Twitter Interactions – a guide to bots, shills and echo chambers

Unsupervised image clustering

UW Alerts Application

VACATION FINDER

VIDEO GAME ANALYSIS

WA State Public School Visualization

Where to Park?

Why is Kramer the best character?

DATA 515 Project Proposal

Monte Carlo LVII

Description

Type: Tool

Jonathan Alexander

The tool will allow users to run monte carlo simulations for various NFL matchups based on player data and past team performance. With the super bowl coming up it will be interesting to see how if we are able to see how accurately we are able to simulate general match-ups and the upcoming super bowl game in particular.

Data Sources (tentative):

Per player stats: <https://www.kaggle.com/datasets/zynicide/nfl-football-player-stats>

Past Super Bowl Matchups: <https://www.kaggle.com/datasets/timoboz/superbowl-history-1967-2020>

Player Participation Data:

<https://www.kaggle.com/datasets/toddsteussie/nfl-play-statistics-dataset-2004-to-present>

Team Stats and Outcomes: <https://www.kaggle.com/datasets/the-devastator/nfl-team-stats-and-outcomes>

Beers, Breweries, and Better Choices (Tool)

Chandler Ault

Pitch:

Have you ever been at the grocery store or brewery and been paralyzed by the number of different beers to choose from? Be paralyzed no longer. My pitch is to combine beer and brewery datasets to create a tool for finding beers and breweries that best match the user's preferences. We can match the beers based on metrics like IBU, astringency, and more! Another feature we could add is a user could input multiple beers they are trying to choose from and our tool would output their best match.

Data Sources:

<https://www.openbrewerydb.org/documentation>

https://www.kaggle.com/datasets/ruthgn/beer-profile-and-ratings-data-set?select=beer_profile_and_ratings.csv

Customer Personality Analysis – Research

Ameya Bhamare

Summary

- Customer Personality Analysis is a detailed analysis of a company's ideal customers. It helps a business to better understand its customers and makes it easier for them to modify products according to the specific needs, behaviors and concerns of different types of customers.
- For example, instead of spending money to market a new product to every customer in the company's database, a company can analyze which customer segment is most likely to buy the product and then market the product only on that segment.

Why do I want to work on this?

- Honestly, I have an ulterior motive. With petabytes of data being generated by most companies, this project would look good on data science resume
- Secondly, there's a plethora of data out there with no bounds on how it can be juiced
- Thirdly, I enjoy understanding how people interact with brands

Data sources

- [Customer Analysis](#)
- [Starbucks Customer Data](#)

DATA515 Project: WA State Public School Visualization

By Matthew Blake

- goal: to create a visualization that captures the academic quality/performance of each school district in wa state. this visualization in factor in metrics such as student growth, student performance, classroom teachers and class size etc.
- project type: visualization tool
- data sources (tentative):
 - Washington Office of Superintendent of Public Instruction (<https://washingtonstatereportcard.ospi.k12.wa.us/>)
 - Washington Geospatial Open Data Portal (https://geo.wa.gov/datasets/23bbd746f9924c149681815cf2a6300_0/explore?location=47.27273%2C-120.802250%2C7.47)

Multilingual OCR for Indic Scripts

~ Soham Butala

Project Type : Research / Tool

A document analysis system in India must support multiple languages at the same time. With emerging multilingualism in urban India, it is often necessary to support bilingual, trilingual, or even more languages. This demands development of a multilingual OCR system which can work seamlessly across Indic scripts.

Dataset :

<https://github.com/ssun32/CLIRMatrix>

<https://github.com/project-anuvaad/anuvaad-parallel-corpus>

Unsupervised image clustering

Zachary Bowyer

Image clustering can be useful for various detection tasks. For this project, we will be combining the power of unsupervised clustering algorithms, principal component analysis, and an interactive visualization to allow users to specify image clustering rules for future grouping (Tool).

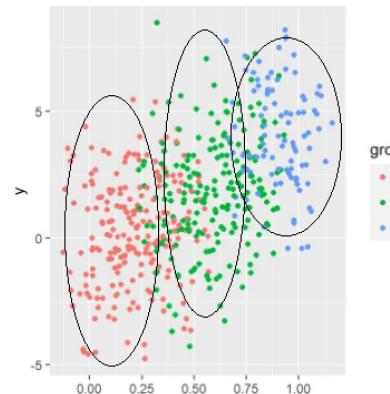
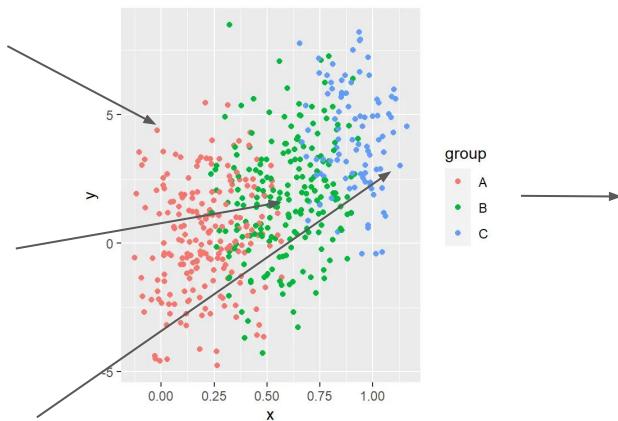
Step 1: Get image batches

Step 2: Run K-means/DBSCAN to get cluster recommendations

Step 3: Plot all images on a 2D plot using PCA, color code based on recommendations

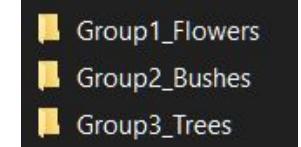
Step 4: Give user ability to select areas of the plot to classify as an image type/group

Step 5: Run new images through user selected groups and output to folders.



Data source 1:
<https://www.kaggle.com/datasets/pavansanagapati/images-dataset>

Data source 2:
<https://cocodataset.org/#home>



Popular Video Games



Marques Chacon

- Project Output: Reusable Data and Tool
 - Creating a bar graph distribution with a time slider to show changes in most common video game genres
 - Building an interactive dashboard that highlights sales, target demographics, and playing time for different games, with dropdown filters for a specific gaming platform
 - Use a machine learning model to predict sales for a particular game, which could be useful for game developers
- Data
 - Two datasets from Kaggle:
 - Data scraped from RAWG (video game database)
 - Video game sales data and critic review scores

CRIME ANALYSIS

Prerit Chaudhary

Project Type : Research

Objective

By examining the data, it's evident that the crime rate has reduced over the last few years. To better understand the cause of this change, I plan to drill down on the underlying factors such as socio-economics and demographics using time-series, spatial, and statistical analysis. Some of the points I would like to explore are:

- What has been the % change in different types of crime?
- What are the specific age groups for victims and suspects?
- Evaluate if there is correlation between economic conditions, such as poverty and unemployment and crime rates.
- Is there any pattern between type of crime and time of the day when it is happening.
- Geospatial or Geographic analysis for each crime and deep dive using Demographic data.

Data Source

[Data.gov](#)

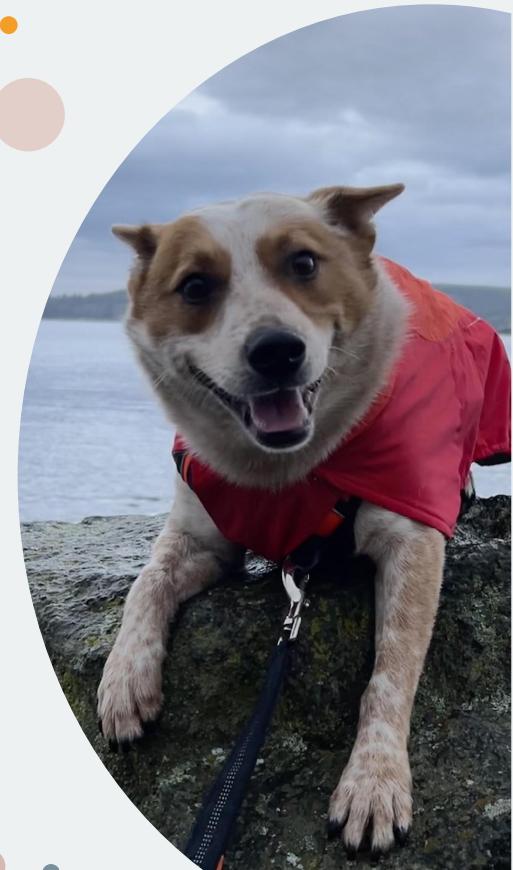
[United States Department of Labor](#)

[Kaggle](#)

Pet Palette

Your Guide to Adoptable Paws Heat Map & Breed Comparison

- Business (or personal) questions of interest:
 - Where can I adopt a dog or cat of a specific breed?
 - What specific breed of dog or cat would be a good fit for my household?
- Project Output: a tool (simple web app)
 - User Selects
 - Animal type
 - Up to 3 (maybe 5?) breeds
 - Location (state dropdown)
 - Visualizations on screen would filter based on the selections
 - Heat map displaying concentrations of adoptable pets of selected breed(s)
 - Comparisons of: Life expectancy, friendliness, shedding, etc
- Data
 - PetFinder API (heat map)
 - Dog Traits: <https://tmfilho.github.io/akcdata/>
 - Cat traits: <https://altef.github.io/logipar/>



STUDY OF GLOBAL PROGRESS TOWARDS SUSTAINABLE DEVELOPMENT GOALS

Name: Anish Dixit

Project Type: Research / Analysis

Project Details:

- In 2015, the UN has defined 13 Sustainable Development Goals targeting different aspects of preservation of life on Earth to be attained by 2030.
- It is important to gauge empirically how each country is working towards attaining the goals.
- Did the Covid 19 pandemic have an effect on countries' attainment of the SDGs?
- Are there geographic/socio-economic factors determining how well a country/region fares in attaining SDGs?
- Is any particular SDG easier to attain than others?

Data Sources: There is a lot of official data available on the UN website which can be used for analysing a variety of aspects, countrywise, goalwise, yearwise, and so much more.

1. <https://unstats.un.org/sdgs/dataportal>
2. <https://unstats-undesa.opendata.arcgis.com/>

Parking Spot Finder –April Gao

Project Type: Tool

Summary: The objective of this tool is to help user to increase their luck on finding street parking. After use input the destination address and arrival time, this tool will use historical paid parking occupancy to find parking spaces that are less likely to be occupied. This tool will also take public event and sport event into account when finding spots.

Datasets:

- Seattle event calendar: <https://www.seattle.gov/event-calendar>
- Paid Parking Occupancy:
<https://data.seattle.gov/Transportation/Paid-Parking-Last-48-Hours-/hiyf-7edq>

Anomaly Detection in Financial Transactions

Project Output:

Tool (A simple web application)

Project Goals:

Develop a model (**unsupervised** or a **semi-supervised**) that can:

- Detect fraudulent credit card transactions, insurance claims, mobile payments or E-commerce
- Identify specific patterns or behaviors to find those fraudulent transactions
- Perform the detection in real-time rather than offline
- Perform detection with good accuracy:
 - Using an ensemble of ML algorithms
 - Balance the trade-off between false positives and false negatives
- Alert relevant stakeholders when a potentially fraudulent transaction is detected.

Data

Sources:

- Credit Card Transactions Dataset (<https://www.kaggle.com/mlg-ulb/creditcardfraud>)
- Car Insurance Dataset (<https://www.kaggle.com/datasets/incarnyx/car-insurance-fraud>)
- Ecommerce Fraud Detection Data
(<https://www.kaggle.com/datasets/aryanrastogi7767/e-commerce-fraud-data>)
- Mobile Transactions (<https://www.kaggle.com/datasets/ealaxi/paysim1>)

Name: Sagnik Ghosal

COVID19: Patient impact and State policy

By Nizan Howard

Type: Tool Project

Objective: To build a tool that can help provide **data mining** and a visual to help with analysis between the two data sets.

Allows comparative visualization between patient and hospital capacity on the state level.

- Preprocess contextual data
- Organizes and scales numeric data
- Provides a visual representation of the preprocess completion

Tentative data sets:

Data set 1: <https://healthdata.gov/dataset/COVID-19-Reported-Patient-Impact-and-Hospital-Capa/6xf2-c3ie>

Data set 2 : <https://healthdata.gov/dataset/COVID-19-State-and-County-Policy-Orders/gyqz-9u7n>

Impact of Sci-Fi among Fictions over the Years

Research / Wanwei Huang

Goals:

1. Join the two datasets to see if Sci-Fis that are generally popular among fictions are popular among Sci-Fis.
2. To see how Sci-Fi evolves over time, what elements are becoming more popular among readers and what are they losing interest in. Could the shift be due to the development of technologies in our time?

Datasets:

1. <https://www.kaggle.com/datasets/cmenca/new-york-times-hardcover-fiction-best-sellers>
2. https://www.kaggle.com/datasets/tanguypledel/science-fiction-books-subgenres?select=sf_time_travel.csv

James Joko

Project Type: Research

I would like to research the effect of different life habits on the quality and efficiency of sleep and sleep patterns (REM, Deep, Awakenings). Different habits I would look into include smoking, alcohol consumption, and exercise frequency/duration.

Data Sources:

- [Sleep Study Dataset](#)
- [FitBit Fitness Tracker Dataset](#)

VIDEO GAME ANALYSIS

Project Type: Reusable Data

Questions:

1. What are the most popular video game genres?
2. Do video games with higher ratings correlate with higher sales?
3. Are games that can be beaten more easily rated higher?
4. Does playtime affect sales?

Data Sources:

RAWG Dataset: Contains 474417 videogames, scrapped using RAWG API.

Video Game Sales: Includes games with sales over 100,000 copies, scrapped from vgchartz.com

Name: Yamina Katariya

Food *Deal*-ivery

Tejal Kolte

Project Output: tool (simple web app)

- Enter your delivery location and desired food order from a specific restaurant
- The app will compare real-time total food delivery charges across different services and provide you with the best *deal!*

Data

- UberEats API
- DoorDash API
- GrubHub API



Title: Kaggle Genie

Project type: Tool

Pitch: Data scientists frequently load data from Kaggle for our modeling projects. This interactive tool loads several Kaggle data sources, and coaches you through data cleaning considerations you'll need to make before training a model.

Data Sources:

- <https://www.kaggle.com/datasets/PromptCloudHQ/us-jobs-on-monstercom>
- <https://www.kaggle.com/datasets/promptcloud/uk-job-listing-dataset-from-monster>

Presented by: Emily Linebarger

Data Analysis: Rotten Tomatoes and the Oscars

Carolina Mack

- Everyone loves to reference Rotten Tomatoes scores, but how accurate and impactful are they really?
- Research Questions:
 - How reliable are Rotten Tomatoes scores as Oscar predictors?
 - What is a more accurate metric for predicting Oscar outcomes: critic score or audience score?
 - Which award type is more closely correlated with Rotten Tomatoes score: best actor/best actress, or best picture?
 - Does the correlation between the two variables change with movie genre?
- Data sources:
 - [Rotten Tomatoes Dataset](#)
 - [Oscars Dataset](#)

Yash Manne: Why is Kramer the best character?

- Project output: Tool (simple web app):
 - **Chronicle** appearances of each character across the series.
 - **Identify correlations** between characters and audience ratings
 - **Search query** for partially remembered dialogue snippets
 - Extensions with increasingly complex NLP models
- Data:
 - Complete scripts for all 173 episodes of Seinfeld ([Kaggle](#))
 - Audience ratings for each episode ([IMDb](#))
 - Viewer Counts ([Wikipedia](#))



Airline Delay Prediction

- Project output: a tool (simple web app)
 - Enter airport and weather conditions (wind, ppt., hazards, etc.)
 - Use a regression or machine learning model to predict the departure delay
- Data
 - Kaggle 2019 airlines delay data
 - NOAA weather data
 - Historical weather data by location
 - NCEI natural disasters data
 - Earthquakes, storms, volcanoes, etc.

Car Decision Making Interactive Helper

Project Type: Web deployed Tool

Aim:

The aim of this project is help anyone interested in buying a car, displaying interactive features/segmentation on a web deployed tool

Data Sources:

<https://www.kaggle.com/datasets/tymekurban/new-cars-usa-202223-dataset>

<https://github.com/abhionlyone/us-car-models-data>

Akshit Miglani

Analyzing the Wisconsin Wolf Population (Research)

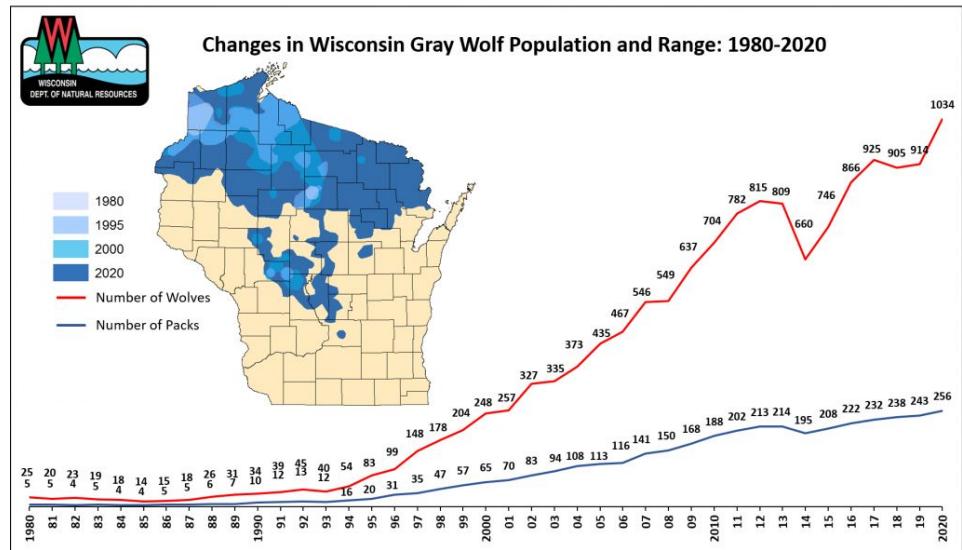
By: Peter Mohammadi

- Over the past 30 years the Wisconsin Wolf population has been rising
- Wisconsin natives say they are dangerous and are causing the deer population to plummet, they want to kill them off
- I want to explore these claims

Data Sources:

<https://dnr.wisconsin.gov/>

https://widnr.widen.net/s/nlrl7tzdw5/wisconsin_gray_wolf_report_2022



Analyzing Academic Sexual Misconduct

Research Project

The ASMD provides a comprehensive dataset of cases of academic sexual misconduct from universities and colleges across the US. This data includes information about the individuals involved, the circumstances of the misconduct, and the outcome of any investigations or proceedings. By leveraging this rich dataset, I want to gain new insights into the possible factors that contribute to academic sexual misconduct, including demographic factors such as age, gender, and ethnicity, as well as institutional factors such as university size, geographic location (extended to political affiliation, education levels of state), and type of institution.

<https://academic-sexual-misconduct-database.org/download-data>

<https://public.opendatasoft.com/explore/dataset/us-colleges-and-universities/table/>

<https://www.pewresearch.org/religion/religious-landscape-study/compare/party-affiliation/by/state/>

Speech To Text Engine

By: Ananya Nair

Project Type: Tool

Description: A speech to text engine is a tool that will allow the user to process human speech into a written format. The objective is to use machine learning in order to integrate grammar, syntax, structure, and composition of audio and voice signals to understand and process human speech.

Datasets:

1. <https://research.google.com/audioset/>
2. <https://www.robots.ox.ac.uk/~vgg/data/voxceleb/>

Recipe Generator

TOOL

A Web App

- Build a **Streamlit WebApp** that generates a recipe
- **Input:** List of ingredients you have at home.
- **Output:** A random recipe that uses only the ingredients listed.

Data Sources

- Recipe Box
- Kaggle

Saumya Nauni

Exploring Weather Trends

Logan O'Brien

- Project Output: Reusable Data (Interactive Dashboard)
 - Provide a tool that allows exploration
 - E.g. is unusually light rainfall in January an outlier or a trend?
- Data Sources:
 - NOAA Data
 - Local Climatological Data
 - Date range up to 1948 – Present
 - Monthly Summary (temperature, precipitation, wind speed, etc.)
 - Washington Department of Ecology
 - Air Quality data for many sites throughout WA
 - Ranging from 2010 - Present



Relation between crime rates and housing prices in Seattle

Aaditya Parthasarathy

- **Project Output:** Research and reusable data (maps and web app)
- The goal of this project is to analyze the data of crime statistics and housing prices in Seattle and identify any relationships between the two factors. The project will involve cleaning and preprocessing the data and creating meaningful visualizations to help illustrate the findings.

- **Data Sources:**

- Housing datasets

- [Kaggle houses sales data](#)

- Crime datasets

- [Opendata API](#)
 - [Crime Stats from 2008 to present](#)



Project Proposal: Create an API Overlay for Smogon.com

- Project Type 2: Create a Tool
- <https://smogon.com> is a website that offers data related to the competitive viability of each Pokemon in the namesake video game series
- It offers information on the stats of each pokemon, what “tier” of competition each Pokemon appears in and is legal for, as well as other information like the list of moves the Pokemon has access to
- The data is publicly available through Smogon’s website but changes quickly as new trends emerge, and is not available in a form that is easy to process in an application
- The plan for this project is to create an API in FastAPI that access a static dataset containing the constant information on each Pokemon and then scrapes the competitive data from Smogon on the fly when it is called for (using BeautifulSoup)
- Datasets:
 - <https://smogon.com>
 - <https://www.kaggle.com/datasets/rounakbanik/pokemon>

Face Extractor Tool

Description

DATA 515 - UW MSDS
Jeremie Poisson

- Tool to extract faces from images using machine learning techniques
- Implementation for training and testing
- Potentially publish as a Python package to pip
- Python packages for:
 - Training utilities given a dataset
 - Testing utilities given a dataset
 - Running utility to extract faces in separate image files given an image
 - Unit testing
- Example commands:

Data Sources

- Datasets (tentative):
 - <https://www.kaggle.com/datasets/dataturks/face-detection-in-images>
 - <https://datagen.tech/blog/face-datasets/>
 - etc.

```
python face_extract.py train <dataset directory> # trains the model
python face_extract.py test <dataset directory> # tests the model
python face_extract.py run <image path> <faces dir> # extract all faces from <image path> into <face dir>
python face_extract.py unit_tests # run unit tests
```

Crime under the Weather

- Project Type: Research and Reusable Data
 - Analyze whether certain weather conditions have an impact on whether more or less crime is committed
 - Have a map to visualize certain streets or areas that could have more/less crime occurring dependent on weather conditions
- Data
 - Seattle gov SPD Crime Data
 - Location of incident
 - Type of incident
 - Weather Data from NOAA
 - Temperature/Precipitation
 - Seattle Specific



Mark Qiao

Term Rate Conversion for Banks

Project Type: Research

Short Pitch:

The bank has various outreach plans to sell term deposits to their customers such as email marketing, advertisements, telephonic marketing, and digital marketing. Telephonic marketing campaigns still remain one of the most effective way to reach out to people. However, they require huge investment as large call centers are hired to actually execute these campaigns. Hence, it is crucial to identify the customers most likely to convert beforehand so that they can be specifically targeted via call.

Data Sources:

1. <https://www.kaggle.com/datasets/prakharrathi25/banking-dataset-marketing-targets>
2. <https://archive.ics.uci.edu/ml/datasets/bank+marketing>

Orphaned Well Mapping Tool

By Emily Rolen

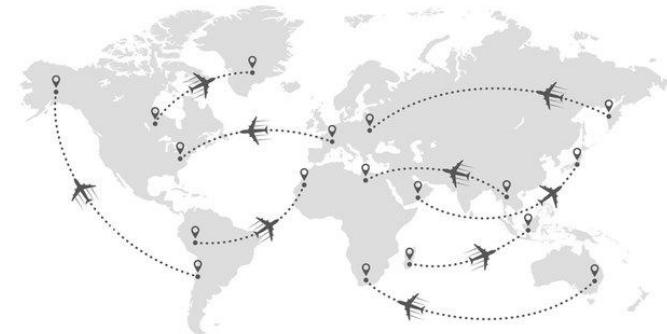
- Proposal:
 - Orphan gas/oil wells present significant health and environmental risks to those that live and work nearby
 - Recently approved federal funding to cap orphan wells has lead to significant data collection efforts
 - No tool currently exists to search orphan well locations
 - Create a searchable map of orphaned gas wells to identify sites near population centers and points of personal interest
- Data sources:
 - [Documented Orphaned Oil and Gas Wells Across the United States](#) (2021 & 2022)
 - [Population Density Information](#) (2010 US Census)



BEAT THE CROWD!

FRANCES SCOTT-WEIS

- Project Output: Web App (Tool)
 - User input: desired destination and/or dates of travel
 - Use a ML model to predict how busy a travel destination will be at some point in the future
 - Output:
 - Time series visualization of predicted busyness over time for a specific destination
 - Heatmap visualization of predicted busyness in different locations for a specific time frame
- Data
 - World Bank International Tourism Demographics
 - United Airlines Traffic and Capacity



Airbnb Recommender – Top 5 Airbnbs in your Desired City in the US

- Project output: A Visualization tool or Web App
 - Recommend Top 5 Airbnb Hosts for a specific US City based on Popularity, Budget, Reviews, Type of Rooms and Neighborhood of your choice
 - Use a machine learning model to classify Airbnbs as High, Medium, Low
 - Create Visualization with filters and recommendations with all listings on the Map
 - Perform overall city and state level # of Airbnbs analysis
- Data
 - Airbnb USA Listings – [link](#)
 - Airbnb US listings for 15 states and 30 unique cities
 - Airbnb Name, Demographics, Review, Price, Room Type
 - US Airbnb Open Data – [link](#)
 - Airbnb Name, Demographics, Review, Price, Room Type



FRAME - Food Recommendations For All Methodical Eaters

Arjun Sharma

Project Type: Tool

Data Sources:

1. [Uber Eats Restaurants and Menus](#)
2. [New York Public Library \(What's on the menu?\)](#)
3. [Data.gov](#)
4. [Yelp Reviews](#)



Zip-Code: 91101

Latitude and Longitude:

Cuisine: American

Food Type: Salad

Calories Per Serving: 400 - 600

Max Distance: 5 mi

Allergies: None

Price Range: \$10 - \$25

Misc. Preferences: High Protein, Low Sodium, Low Fat, Vegan

Restaurant Rating: 3+

Brussels Sprouts

Magnolia House - ★★★★☆,
336 S Lake Ave Pasadena, CA
91101,
2.4 mi away,
\$14.99,
450 Calories

Caesar Salad

GRANVILLE - ★★★★,
270 S Lake Ave Pasadena, CA
91101,
3.4 mi away,
\$17.99,
550 Calories

Buddha Bowl

Urban Plates - ★★★★☆☆,
269 S Lake Ave Pasadena, CA
91101,
1.8 mi away,
\$21.99,
570 Calories

Procedure:

1. User preferences taken as input.
2. Based on Content-Based Filtering and Collaborative Filtering, suggest the top 10 dishes and the restaurants that offer them.
3. Show relevant details on clicking on a suggested dish (Restaurant address, Distance from location, Nutritional Information, Price).
4. Show the 5 most recent reviews of the restaurant.

Global Power Plants and Energy Consumption Analysis

Mithali Shashidhar

Project Output: research and reusable data (data visualization and analysis)

- Research how economic indicators affects a country's energy choices.
 - Do countries that have a higher GDP invest in clean energy more than others?
 - Are countries with a higher proportion of educated youth per capita more inclined towards using renewable sources of energy?
- Build map visualizations that answer these questions.
- Include search bars and filters to drill down on countries.
- Data:
 - <https://www.kaggle.com/datasets/ramjasmaurya/global-powerplants>
 - <https://databank.worldbank.org/source/world-development-indicators>

Airline Crew Pairing/Scheduling (Research)

Problem

Update now-antiquated current industry solutions for optimizing scheduling flights for commercial airlines.

Objective: Minimize costs (airplane expenses, fuel, crew, and equipment) while ensuring all flights and crew members are covered exactly once.

Crew Pairing → Assign airline crew members to a crew such that all crew members' constraints are satisfied

- Max. hrs/week
- Requested vacation time
- Domestic/International rules

Crew Scheduling → Assign crews to a series of flight such that all FDA/safety constraints are satisfied

- Flight connectivity
- Efficient route

Datasets

<https://www.sciencedirect.com/science/article/pii/S2192437620300820>

<https://github.com/crewml/crewml>

<https://github.com/SreekarJammula/Flight-Data-Analytics->

Twitter Interactions – a guide to bots, shills and echo chambers

Raman SV

Project type: Create a Tool

Data sources :

- 1) Twitter API via their developer platform
- 2) Google News API for trending topics

Twitter Interactions – a guide to bots, shills and echo chambers

Keywords - Twitter, Crypto, Tesla

Multiple selections

Number 1

Elon Musk @elonmusk
179 Following
128.2M Followers

Number 2

Crypto.com @cryptocom
The best place to buy, sell, and pay with crypto #BTC #CRO #DeFi #FTTB
435 Following
2.6M Followers

Number 3

@toocryptotoquit
You know, crypto and stuff...live and let live...an equal opportunity critic...demonstrably naive...opinions are transitory...just be nice
697 Following
413 Followers

Number 4

Vinny D'Agostino @proofofwork1
Husband and father to 3 experiments. Former FBI Agent (NY), lawyer, comic and cyber investigator. #bitcoin, not crypto
638 Following
3,006 Followers

Most Tweeted Words

- 1) Tesla
- 2) To the moon
- 3) Free Speech
- 4) Twitter
- 5) Inflation

Most Common Followers Bio

- 1) Crypto
- 2) Free Speech
- 3) Echo Chamber
- 4) #Twitter
- 5) #MuskFan

Similar Accounts

- 1) @POTUS
- 2) Not Jerome Powell
- 3) WallStreetBets
- 4) CryptoBros
- 5) @POTUS

Pitch

- 1) Search by key terms and get results of top Twitter users for these terms (KNN Algorithm?)
- 2) Click on the results and get a deep dive on bot presence and retweets for this profile
- 3) View similar profiles and their bios, which resets the list to show their top followers and catch words
- 4) Integrate trending news and hotlist of words to search for users who populate these echo chambers

Charge Me Up!

- Project output: research & map visualizations
 - Research electric vehicle charges to check correlation between charging stations and vehicle registration
 - Build map visualizations to show electric vehicle and charger density for various types of vehicles, electric vehicle range, and chargers
- Data
 - Electric vehicle (battery and plug-in hybrid) vehicle registry in Washington State [data.wa.gov]
 - Elective vehicle charging station locations [afdc.energy.gov]



VACATION FINDER

A TOOL FOR PREDICTING POPULAR TRAVEL DESTINATIONS

ADITHYAA V

- PROVIDE INSIGHTS INTO THE DESTINATIONS THAT ARE LIKELY TO BE THE MOST POPULAR IN THE FUTURE, HELPING TRAVELERS PLAN THEIR TRIPS.
- PROVIDE VALUABLE INFORMATION FOR GOVERNMENTS AND TOURISM ORGANIZATIONS TO HELP THEM PLAN FOR THE FUTURE AND MAKE DECISIONS ABOUT HOW TO ATTRACT MORE TRAVELERS.
- TRAVELERS WILL BE ABLE TO PLAN THEIR TRIPS WITH CONFIDENCE, KNOWING THAT THEY WILL BE VISITING THE DESTINATIONS THAT ARE LIKELY TO BE THE MOST POPULAR IN THE FUTURE.
- PREDICT THE MOST POPULAR TRAVEL DESTINATIONS BASED ON THE HAPPINESS INDEX, THE NUMBER OF INTERNATIONAL TRAVELERS, ETC.
- UTILIZE MACHINE LEARNING ALGORITHMS, SUCH AS LINEAR REGRESSION, DECISION TREES, AND RANDOM FOREST, TO GENERATE PREDICTIONS.

DATA SOURCES:

1. WORLD BANK DATA ([HTTPS://DATA.WORLDBANK.ORG/INDICATOR/ST.INT.ARVL](https://data.worldbank.org/indicator/st.int.arvl))
2. WORLD HAPPINESS REPORT ([HTTPS://WORLDHAPPINESS.REPORT/ED/2020/](https://worldhappiness.report/ed/2020/))

Climate Change and Wildfire in the US: An Analysis

Project Type: Research Project

Aim:

The aim of this project is to investigate the effect of climate change on the likelihood and severity of wildfires in the United States. The goal is to understand the relationship between wildfires and climate factors by combining data on wildfires and climate factors with information on land cover and vegetation type. My goal is to provide insights and recommendations for fire prevention and management strategies in a changing climate.

Data Sources:

[National Interagency Fire Center \(NIFC\) – Science Base](#)

[United States Geological Survey \(USGS\) – Science Base](#)

[National Centers for Environmental Information \(NCEI\)](#)

Ishank Vasania

Electric Vehicle Charging Heat Map

- Project Output: a tool that helps visualize the popularity of a charging station for electric vehicle (EV) charging
 - Calculate a popularity score for EV charging stations based on zip code, capacity, number of nearby charging stations and number of EVs registered in the area
 - Create heat map based on charging station locations
 - The purpose of the tool is to indicate if a certain area is lacking charging stations or if any existing charging stations will likely to be full
- Data
 - Washington state EV population data ([Link](#))
 - Washington state EV charging stations data ([Link](#))
- Contact: Jason Wang

Where to Park?

A tool to help people find the best place to park their car around destination.
Jenny Wong

Data Sources:

1. Seattle GeoData Blockface API

<https://data-seattlecitygis.opendata.arcgis.com/datasets/SeattleCityGIS::blockface/about>

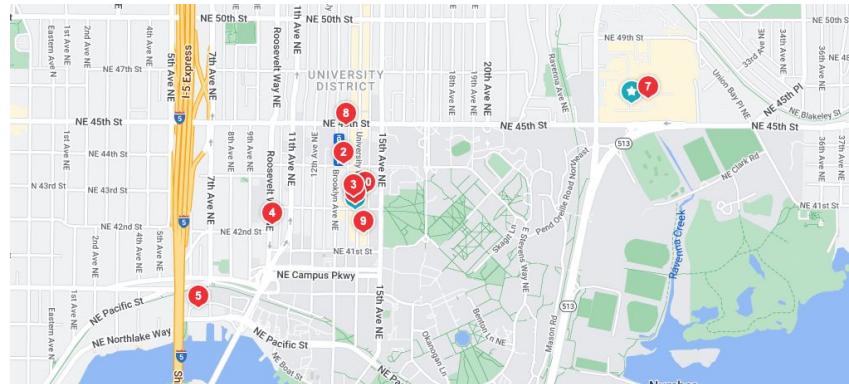
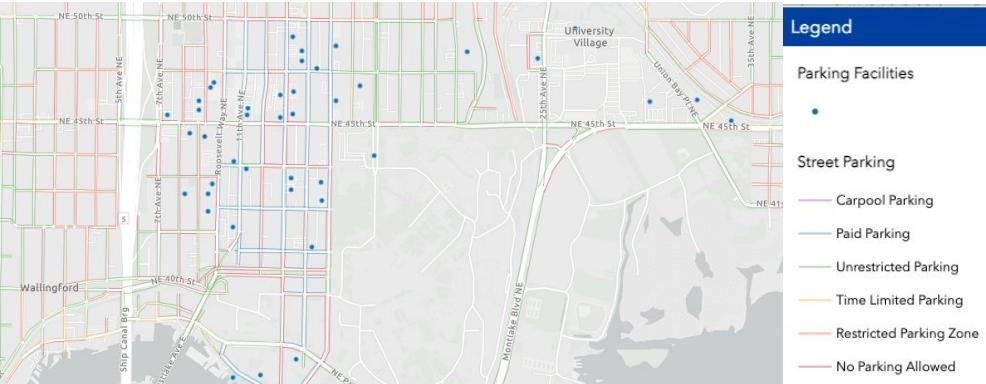
1. Yelp Fusion API

<https://docs.developer.yelp.com/docs/fusion-intro>

Main Idea:

Seattle GeoData developed a map labeling every street for whether it is able to street parking or not. However, it is not searchable by names of the places.

I would like to combine the **location of stores** in Yelp with the **parking map**, to give users a list of possible street parking location or garage, and rank and recommend the best location.



Tourist Places Recommendation System

Name: Tanushree Yandra

Project Type: Tool

About:

The aim of this tool is to recommend popular places in and around Seattle visited by tourists based on factors like the type of place the user wants to visit (museums, cafes, trails, beaches, etc.), proximity, budget, day, time, weather, and so on. I have always found recommendation systems interesting and thought it would be really cool to build one platform that suggests places on such a large number of factors.

Datasets:

<https://mygeodata.cloud/data/download/osm/tourist-attractions/united-states-of-america--washington/king-county/seattle>

<https://www.kaggle.com/datasets/oklena/seattle-area-restaurants>

<https://data.seattle.gov/Parks-and-Recreation/Seattle-Parks-and-Recreation-GIS-Map-Layer-Web-Ser/f5g8-sb93>

Seattle Weather Prediction

- Project output: research & reusable data (map visualizations)
 - Build a map visualization to show how temperature and precipitation change in Seattle
 - Research how season, temperature and precipitation are related
 - Predict whether it will rain on a certain day based on temperature and precipitation on previous days
- Data:
 - NOAA data of weather
 - Temperature and precipitation dataset from Kaggle

UW Alerts Application

Evan Yip

Project type: Tool Project

Project Pitch: The goal of this project is to build a interactive web application and visualization that displays the trends and updates from the UW alert notification system. The goal is to be able to create a proof of concept visualization that parses through real-time UW alerts onto a map. This is a similar application to the citizen app.

Tools: Plotly, Mapbox, Flask

Data Sources:

1. [UW Alerts Blog](#)
2. [Seattle GeoData Streets](#)



January 26, 2023

UW Alert Seattle

UPDATED at 2:39 p.m. Thursday: Police have not found the man reportedly running near NE 47th St. and 16th Ave. NE with a gun, but continue to patrol the area. The man is described as between 40 and 50 years old, wearing a black, puffy jacket. The area is considered to be reopened. Please stay vigilant and do not engage with anyone who may be armed. Report any potential sightings to 911.

