PROJECT OVERVIEW:  **De-Mystifying ML** Project

Team Members: SEAN McCray, **Reema Singh**, **Jessika Virtanen**,

MELLISA Wood, SERENA TERRY, kimberly gordon

2/5/2019

## Topic for review – Disney Wait Time Data for Machine Learning and Data Science Projects

TouringPlans and the Unofficial Guide book series have their roots in academia. To support the next generation of theme park researchers, we're happy to provide access to our own sets of Disney wait times. Each dataset is a comma-separated file (CSV) of ASCII text, updated monthly. Each row is one discrete observation. Besides the wait times, we were provided a file of features common to each wait time. Each column is a feature captured along with the wait time (e.g., day of week, time of day).

* Sample of Data Sets being used
  + <https://touringplans.com/walt-disney-world/crowd-calendar#DataSets>
  + <https://touringplans.com/walt-disney-world/historical-crowds>
  + <https://disneyworld.disney.go.com/resorts/map/>
  + https://disneyworld.disney.go.com/en\_GB/faq/bands-cards/understanding-magic-band/
* Observation on Disney Properties: Such as Magic Kingdom and Epcot
* Analytics considering:
  + Use of Disney Crowd Calendar, Magic Band, etc. Does it help you avoid the lines?
  + Distance between rides, events and food. Do these items help maximize your time in visiting more items during a day?
  + Time of year. Are the lines shorter/longer during a certain time of year?
* Expected Outcome – There is a way to minimize the impact of waiting in lines while visiting the park even during peak times.

## Project Scope

* For the team to find manageable files to use ML in the context of technologies learned
* Very quick project with less than two weeks to review and work with team.
* Prepare a 15 minute “data deep dive” or “infrastructure walkthrough” that shows machine learning.

## Approach

* Team decided on a problem worth solving, analyzing and visualizing.
* Researched the availability of data.
* Analyzed possible information that can be used as review, aggravation, etc.
* Performed various analysis on the data to analysis and visualization.

## Finding Data

* Resources used to find the appropriate data considered and included
  + APIs
  + Data scraping
  + Large Data Sets from:
    - Disney
    - Parks and Recreation sites
    - Travel sites

## Details of Analytics to include

* File Names

|  |  |
| --- | --- |
| File Name | Summary of Data |
| https://cdn.touringplans.com/datasets/kilimanjaro\_safaris.csv | attraction wait times |
| https://cdn.touringplans.com/datasets/expedition\_everest.csv | attraction wait times |
| https://cdn.touringplans.com/datasets/dinosaur.csv | attraction wait times |
| https://cdn.touringplans.com/datasets/spaceship\_earth.csv | attraction wait times |
| https://cdn.touringplans.com/datasets/soarin.csv | attraction wait times |
| https://cdn.touringplans.com/datasets/rock\_n\_rollercoaster.csv | attraction wait times |
| https://cdn.touringplans.com/datasets/toy\_story\_mania.csv | attraction wait times |
| https://cdn.touringplans.com/datasets/splash\_mountain.csv | attraction wait times |
| https://cdn.touringplans.com/datasets/pirates\_of\_caribbean.csv | attraction wait times |
| https://cdn.touringplans.com/datasets/navi\_river.csv | attraction wait times |
| https://cdn.touringplans.com/datasets/flight\_of\_passage.csv | attraction wait times |
| https://cdn.touringplans.com/datasets/slinky\_dog.csv | attraction wait times |
| https://cdn.touringplans.com/datasets/alien\_saucers.csv | attraction wait times |
| https://cdn.touringplans.com/datasets/7\_dwarfs\_train.csv | attraction wait times |

* Summary of type of Data

|  |  |  |
| --- | --- | --- |
| **Variable\_Name** | **Variable\_Description** | **Format** |
| DATE | Park Day (not actual date stamp of the wait time, since some are after midnight) | MM/DD/YYY |
| DATETIME | date-time stamp of wait time | YYYY-MM-DD HH:MM:SS |
| SPOSTMIN | Standby Posted Wait Time (in minutes) \* | numeric |
| SACTMIN | Actual Wait Time (in minutes) | numeric |

* File format (fixed length, comma separated, etc.)
* Types of data manipulation used for analysis
  + Python: Pandas, Matplotlib, Tweepy and Flask
  + HTML/CSS/Bootstrap
  + JavaScript Plotly and D3.js
  + JavaScript Leaflet
  + MySQL Database
  + MongoDB Database
  + Google Cloud SQL
  + Amazon AWS
  + Tableau

## Resources

* Average Wait Time (Min and Max) for Federal Holidays (Each Park) – Melissa
* Interactive Tableau Map with names of Rides – Serena & Sean
* ML & AI – Reena & Kimberly
* Website - Jessica