

Lauren Deans

CSPB 4122

Project Dataset Description

**Data Quality and Cleaning:** The core dataset for this project was obtained from <https://www.kaggle.com/mikeshout/14erpeaks/version/4> . The dataset was downloaded as a CSV file containing 58 rows and 16 columns. This dataset includes data for each peak such as elevation, mountain range, latitude, longitude, standard route, distance, elevation gain, difficulty, traffic low, and traffic high. As far as I am aware, the data included in the main dataset is clean and without erroneous entries. This core dataset was also extended to include data for mapping the elevation profile and standard routes for each peak. A csv file for each 14er was downloaded as a gpx file and converted to csv format. The data for each peak includes standard route information, as well as the elevation profiles along each route. The elevation profile/ route files for each peak will need to be cleaned as there are empty columns for 'name' 'desc' and 'type' attributes. There is also a row in the middle of each file that contains the column names again, which will also need to be cleaned. The dataset will also be enriched to include weather data, which I have already gained access to via DarkSky API . This weather data will be refreshed during use of the visualization and is based on latitude and longitude coordinates, so no weather data is included with this assignment.

**Deriving Attributes and/or Integrating Multiple Datasets:** To enrich the main dataset, I obtained data for each peak which elevation profiles and latitude and longitudinal coordinates for the standard route to climb each peak. I will also be deriving weather attributes from the latitude and longitude coordinates for the top of each peak. This data will be updated upon interaction with the visualization (dynamic data), but I have already started the coding to obtain current weather information from the coordinates (See Figure 1). Separate from the main dataset are the csv files containing elevation profiles for each peak, along with route information. This data will be used to map the standard route on the interactive map, in addition to comparing the elevation profiles of chosen peaks.

As it is currently, the dataset provides adequate information to carry out the tasks defined in the project description. Browsing the interactive map of Colorado using characteristics such as mountain range, difficulty, distance, popularity, or elevation profile can be accomplished using the main dataset in addition to the individual datasets for each peak. Locating specific peaks and checking current weather at the top of each peak can be accomplished using the latitude and longitude coordinates for each peak, and standard routes can be displayed on the visualization using the individual file for each peak that contains elevation profile and standard route information. Comparing characteristics such as elevation profiles, current

weather, foot traffic, difficulty, and distance can be accomplished using both the main dataset in addition to the files for each individual peak.

```
j> from datetime import datetime as dt
import pandas as pd
from darksky import forecast
from datetime import date, timedelta

key = '9bcefc43dbdca5eac36ceab97a114ccb'

df = pd.read_csv('14er.csv')
MountainPeakInfo = df[['Mountain_Peak', 'Lat', 'Long']]
Elbert = MountainPeakInfo.loc[0].values
NamePeak = Elbert[0]
Lat = Elbert[1]
Long = Elbert[2]

Peak = key, Lat, Long
now = dt.now().replace(microsecond=0).isoformat()
peakweather = forecast(*Peak, time=now)
temp = peakweather['currently']['temperature']
wind = peakweather['currently']['windSpeed']
summary = peakweather['currently']['summary']
precipProb = peakweather['currently']['precipProbability']
precipIntensity = peakweather['currently']['precipIntensity']
weather_string = ("Current Weather for {} \n" +
" Temperature: {} F \n" +
" Wind Speed: {} mph \n" +
" Weather Summary: {} \n" +
" Precipitation Probability: {} \n" +
" Precipitation Intensity: {} \n")
print (weather_string.format(NamePeak, temp, wind, summary, precipProb, precipIntensity))

Current Weather for Mount Elbert
Temperature: 42.39 F
Wind Speed: 19.58 mph
Weather Summary: Windy
Precipitation Probability: 0
Precipitation Intensity: 0
```

Figure 1. Example of obtaining current weather information for Mount Elbert

**Transformation in Proper Format:** The main dataset containing information on each mountain peak is clean and already in csv format. I also located files for each peak that contain elevation profiles along the standard route to climb each peak, and each of those were manually converted to csv format from gpx format. These files were downloaded individually from 14ers.com and converted via an online converter.

**Data Abstraction:** Below are the first two rows from the main dataset as well as the first two rows from one of the peak's elevation files, followed by data abstraction for each dataset using Munzner Ch. 2 framework

Mountain_Peak	Mountain_Range	Elevation_ft	fourteener	Prominence_ft	Isolation_mi	Lat	Long	Standard_Route	Distance_mi	Elevation_Gain_ft	Difficulty	Traffic_Low	Traffic_High
Mount Elbert	Sawatch Range	14440	Y	9093	670.00	39.1178	-106.4454	Northeast Ridge	9.50	4700	Class 1	20000	25000
Mount Massive	Sawatch Range	14428	Y	1961	5.06	39.1875	-106.4757	East Slopes	14.50	4500	Class 2	7000	10000

	latitude	longitude	altitude (ft)
1	39.151771069	-106.419180632	10061.0
2	39.187499285	-106.475614309	14357.9
3	39.151817322	-106.419113159	10061.0

Column Name	Semantics	Data Type	Attribute Type	Availability
Mountain_Peak	The name of the peak	String	Categorical - nominal	static
Mountain_Range	The name of the primary mountain range the peak is apart of	String	Categorical - nominal	static
Elevation_ft	The peak elevation in feet	Numeric	Quantitative - sequential	static
Fourteener	Indicator if the peak is considered a 14er, Y or N value	Boolean	Ordinal	static
Prominence_ft	How much higher peak is in feet from the next highest point	Numeric	Quantitative - sequential	static
Isolation_mi	Distance in miles from the nearest point with same or higher elevation	Numeric	Quantitative - sequential	static
Lat	Latitudinal coordinates in decimal form	Numeric	Positional	static
Long	Longitudinal coordinates in decimal form	Numeric	Positional	static
Standard_Route	Name of most commonly used hiking route to peak	String	Categorical - nominal	static
Distance_mi	Distance of standard route in miles	Numeric	Quantitative - sequential	static
Elevation_gain_ft	Elevation gain of standard route in feet	Numeric	Quantitative - sequential	static
Difficulty	Yosemite Decimal System, a value ranging from Class 1 (easiest) to Class 5 (most difficult)	String	Ordinal	static
Traffic_Low	Low range of estimated visits in 2017	Numeric	Quantitative - sequential	static
Traffic_High	High range of estimated visits in 2017	Numeric	Quantitative - sequential	static

Column Name	Semantics	Data Type	Attribute Type	Availability
Latitude	Latitudinal coordinates in decimal form	Numeric	Quantitative - sequential	static
Longitude	Longitudinal coordinates in decimal form	Numeric	Quantitative - sequential	static
Altitude(ft)	The elevation in feet	Numeric	Quantitative - sequential	static