(https://colab.research.google.com/github/presidentlines/AvalancheVol3/blob/main/everettsAvalanch

Team AVYULAUNCH

E R N K L

v o a a e

e c t n e

r k h e

e f a

t o n

t r

d

11/13/2021

Combine weather and avalanche data!

In [1]:
```
import pandas as pd
import numpy as np
```

In [2]:
```
pd.set_option('display.max_rows', 20)

df_weather = pd.read_csv('weather.csv')
df_weather = df_weather[['NAME', 'DATE', 'PRCP', 'SNWD', 'TMAX', 'TMIN']]

# Get month dummy variables
df_weather['DATE'] = pd.to_datetime(df_weather['DATE'])
df_weather['MONTH'] = pd.DatetimeIndex(df_weather['DATE']).month
df_weather['MONTH'] = df_weather['MONTH'].astype(str)
df_weather = pd.get_dummies(df_weather, columns=['MONTH'], drop_first=True)

def assign_region(name):
    if name == "BEN LOMOND PEAK, UT US": return "Ogden"
    if name == "ALTA, UT US": return "Salt Lake"
    if name == "BEN LOMOND TRAIL, UT US": return "Ogden"
    if name == "MONTE CRISTO, UT US": return "Logan"
    if name == "BUES CANYON UTAH, UT US": return "Ogden"
    if name == "RAY S VALLEY UTAH, UT US": return "Uintas"
    if name == "SNOWBIRD, UT US": return "Salt Lake"

df_list = []
# Create snow difference by NAME
by_location = df_weather.groupby('NAME')
for name, group in by_location:
    group['Region'] = assign_region(name)
    group['snow_diff'] = group['SNWD'] - group['SNWD'].shift(1)
    group['temp_range'] = group['TMAX'] - group['TMIN']
    df_list.append(group)

df_weather = pd.concat(df_list)
print(len(df_weather))
```

```
63405
```

In [5]:
```
# Clean up avalanche data
df_avalanche = pd.read_csv('avalanches-11-13-2021.csv')
df_avalanche = df_avalanche[['Date', 'Region']].dropna()
df_avalanche['DATE'] = pd.to_datetime(df_avalanche['Date'])
df_avalanche['Ahvyulaunsh'] = 1
```

In [6]:
```
# Combine avalanche and weather on region and date
df_combined = pd.merge(df_weather, df_avalanche,  how='left', on=['Region',
```

In [7]:
```
# Create variable of summed up avalanches per day
summed = df_combined.groupby(['NAME', 'DATE'])['Ahvyulaunsh'].agg('sum').re
summed['Avalanche'] = summed['Ahvyulaunsh']
reassembled = pd.merge(summed, df_combined, how='left', on=['NAME', 'DATE']
reassembled = reassembled.drop_duplicates()

# Final clean up
reassembled['Avalanche_binary'] = reassembled['Avalanche'] > 0
reassembled = pd.get_dummies(reassembled, columns=['Region'], drop_first=Tr
reassembled = reassembled.drop(['Ahvyulaunsh_x', 'Ahvyulaunsh_y', 'Date', '

df_final = reassembled.dropna()
df_final.to_csv("combined_weather_avalanche.csv")
#print(df_final.columns())
```

In [9]:
```
df_final
```

Out[9]:

	DATE	Avalanche	PRCP	SNWD	TMAX	TMIN	MONTH_10	MONTH_11	MONTH_12	MONTI
1	2000-01-02	0.0	0.64	43.0	19.0	4.0	0	0	0	
2	2000-01-03	0.0	0.54	51.0	13.0	6.0	0	0	0	
3	2000-01-04	0.0	0.11	52.0	28.0	-1.0	0	0	0	
4	2000-01-05	0.0	0.93	66.0	22.0	11.0	0	0	0	
5	2000-01-06	0.0	0.00	62.0	23.0	3.0	0	0	0	
...	...	...	...	...	...	...	...	...	...	
68545	2021-11-03	0.0	0.00	10.0	44.0	31.0	0	1	0	
68546	2021-11-04	0.0	0.00	10.0	53.0	40.0	0	1	0	
68547	2021-11-05	0.0	0.00	9.0	51.0	39.0	0	1	0	
68548	2021-11-06	0.0	0.00	8.0	48.0	43.0	0	1	0	
68549	2021-11-07	0.0	0.00	8.0	43.0	31.0	0	1	0	

41007 rows × 23 columns

In [ ]:
```
```