**Ski Resort Reviews and Resort Statistics**

**Python – Andrew Boral**

This project explores two datasets to find out more about ski centers in the United States. All the project files can be found at the following github repository:

<https://github.com/joeskier2020/ski-resorts>

The github repository includes three Jupyter Notebook files containing the analysis as well as the two Kaggle dataset which were utilized in the analysis. NLTK and Gensim will need to be installed to run the code. Directions are provided inside the Jupyter notebook file. And all the python has been tested to run without errors if the libraries are installed. Some code and pre-processing take a couple of minutes to run.

**Ski Center Reviews:**

Data Source: <https://www.kaggle.com/datasets/fredkellner/onthesnow-ski-area-reviews>

Data File: OnTheSnow\_SkiAreaReviews.csv

My Python Code to Analyze Source:

1. FinalProject NLTK By Review.ipynb
2. FinalProject NLTK By Ski Center.ipynb

Description:

The first dataset is composed over 16,000 text reviews. Natural Language Processing tools were used to analyze the reviews. The data was preprocessed to remove capitalization, remove punction and numbers, and remove common English stop words. Next, TF-IDF analysis was performed to reduce the weight of stop words common to this corpus.

The python libraries utilized to complete the analysis were NLTK and Gensim.

**Ski Center Statistics:**

Data Source: <https://www.kaggle.com/code/beaubellamy/ski-resorts>

Data File: skiResort.xlsx

My Python Code to Analyze Source:

1. FinalProject-SkiResortStats.ipynb

Description:

The second data set includes statistics on ski resorts from around the world. Numerous visualizations are provided. Some important statistics which are analyzed are:

1. Variety of ski slopes
2. Size of ski resorts
3. Difficulty of ski trails
4. Quality of ski lifts.
5. Filtering ski centers without chairlifts