CWRU Data Analytics Boot Camp

Spring 2020

Project Proposal

Chris Bock, Sumeet Dhawan, Kafui Ahedor, Suraj Thyagaraj

PyWeatherman

Background

One of the pivotal topics of discussion in the 2020 Election Debate Rounds is the candidate’s stance on policy changes regarding Climate Change and Global Warming. Worldover, there is increasing awareness and consensus regarding climate change and the immediate need to put policies into action. With this project, we hope to gather enough preliminary data to prove that Global Warming is real and convince the administration to take sweeping action to mitigate the effects of global warming.

Proposed Project

Using Python programming and Jupyter Notebook to gather historic weather data from close to fifteen cities spread across continental United States and summarize the weather trends over the past several decades. By visualizing and analyzing the weather trends, we hope to be able to conclusively say if in fact, Earth has been warming up. We also aim to correlate weather patterns with important human milestones (industrial revolution, automobile industry expansion) with weather related natural disaster frequencies (wildfires, hurricanes, tsunamis, etc.).

Steps:

1. Defining scope of the project

Because of the broad and open-ended nature of the topic in question, our group started off with a very wide focus and based on our assessment of the time constraints over the initial days, decided to narrow it down further. Presently, the goal of our study is to do an analysis of the historical weather of 12 cities in the continental US (3 in each time zone) and to conclusively say if the trends in weather data indicate an overall rise in temperatures across the board.

1. Identifying data source

After trying many sources, we decided to source Weather data from the ‘meteostat’ weather repository (<https://api.meteostat.net/>) using API calls. The JSON object returns are easily accessible for further analysis.

1. Data access and retrieval

Meteostat provides free, unlimited API calls to their repository. After requesting a key, data access was established using information provided in the getting started guides from meteostat,

1. Cleaning up data

4a): Same sample across levels

4b) Missing data

4c) Noisy data-outliers

4d) Binning/Structuring data

4e) Data Transformation

1. Analyzing data

5a) Preliminary visualization: scatter plot

5b) Trends

5c) Regression fits

1. Data Visualizations

6a) Individual Trends-Subplots

6b) Gmaps overlays

6c) Heatmaps

6d) Video timelines

1. Preliminary Statistics

For performing the statistical analysis to state with confidence that the temperatures have significantly risen over the time period that we are investigating, we will be doing a **paired t-test** between the following two data frames:

1. @ time 0 [1951]: quarterly temperature info at 12 cities [1x48]
2. @ time 1 [2016]: quarterly temperature info at 12 cities [1x48]

H0: There exists no significant difference between the means of the paired observations.

H1: There exists a significant difference between the means of the paired observations.

For testing significance in the analysis, an alpha of 0.05 is used as the cutoff for significance. If the p-value is less than 0.05, we will reject the null hypothesis that there's no difference between the means and conclude that a significant difference does exist. If the p-value is larger than 0.05, we cannot conclude that a significant difference exists.

1. Summarize Findings
2. Prepare presentation
3. Final Presentation

**Gantt Chart**

****