Investigating 10 Years of Charlottesville Weather

2nd Year Data Science Case Study by Sarah Menchavez

Background:
As the world continues to face the alarming realities of climate change, what are the effects we observe in our own backyard? Environmental scientists at the University of Virginia have been tracking Charlottesville weather data for the past ten years to examine trends that may inform us about the future of climate change in the region. What conclusions can we draw from this data to help scientists and activists address Charlottesville's most pressing climate concerns?

Deliverable:
Your task is to preprocess the provided data for analysis, use exploratory data analysis (EDA), and perform statistical analysis testing to determine if there are significant differences in weather patterns between different years or seasons. Exploratory data visualization and statistical analysis decisions can be made at your discretion (i.e. you can choose to visualize the distribution of precipitation throughout a year and calculate the average precipitation over ten years).

Your results should be in a presentation format including at least one representation of data visualized and findings from at least one statistical analysis test. In addition to the presentation, submit to Canvas a link to your Github repository including all materials used in the analysis.

All presentations will be made to UVA environmental science researchers at the end of the course.

Charlottesville Weather – Analysis Rubric

DS 4002 - Spring 2024
Due: Wednesday, May 10
Submission format: GitHub repository (submitted by link to canvas)

Individual Assignment

General Description: Submit to Canvas a link to your GitHub repository for this project.

Preparatory Assignments – CS2

Why am I doing this? The goal of this assignment is to practice data preprocessing, visualization, and statistical analysis testing, and to understand the application of data science in real world scenarios such as climate change. By the end of this assignment you will be able to draw conclusions from data with evidence to support your claims.

- Course Learning Objective: create a presentation to deliver to environmental scientists explaining trends and patterns in Charlottesville weather data

What am I going to do? You will begin by understanding the importance of data in climate science research from the background document. With the provided information and dataset you will perform data preprocessing, visualization, and statistical analysis testing to investigate weather patterns. You will then format your findings into a presentation to deliver to environmental scientists at the University of Virginia.

Tips for success:

- Take your time on data preprocessing; make decisions on how dataset formatting can be changed for you to better understand the information
- Try different initial exploratory data analysis approaches; because the data can be broken down into year, season, month etc. think of different and interesting ways this information can be isolated and translated into various types of visual representations
- Research different statistical analysis techniques; you may find unfamiliar but more helpful models
- Remember the scope of your dataset; when drawing conclusions, bear in mind that this data only represents a singular region of Virginia and only the past 10 years
- Present efficiently; highlight your research question, modeling approach, and findings–including only the most important information
- Have fun! It is ok to face challenges; not all analyses will be completed perfectly

How will I know I have succeeded? You will meet expectations on this assignment when you follow the criteria in the rubric below.

Formatting	● One Github Repository (submitted via link on canvas) ● The repository will adapt parts of the TIER Protocol 4.0. In a nutshell, the top level page of the repository should contain: ○ A README.md file (which auto displays) ○ A LICENSE.md file (use MIT as default) ○ A SCRIPTS folder ○ A DATA folder ○ AN OUTPUT folder o A PRESENTATION folder
README.md	● Goal: This file serves as an orientation to everyone who comes to your repository, it should enable them to get their bearings. ● Use markdown headers to divide content. ● Make an H2 (##) section explaining the contents of the repository ● Section 1: Software and platform section ○ The type(s) of software you used for the project. ○ The names of any add-on packages that need to be installed with the software. ○ The platform (e.g., Windows, Mac, or Linux) you used. ● Section 2: A Map of your documentation. In this section, you should provide an outline or tree illustrating the hierarchy of folders and subfolders contained in your Project Folder, and listing the files stored in each folder or subfolder. ● SCRIPTS folder ○ Links to all code produced (i.e. Jupyter notebook) ● DATA folder ○ Links to all datasets used and produced ● OUTPUT folder ○ Images of all charts and figures produced with summaries of findings ● PRESENTATION folder ○ Your final presentation deliverable
LICENSE.md	● Goal: This file explains to a visitor the terms under which they may use and cite your repository. ● Usually, the MIT license is appropriate.
SCRIPTS folder	● Goal: This folder contains all the source code for your project. ● Include all the scripts you used. Try to name each script according to the order it needs to be executed to reproduce the results.

	● Include copious comments explaining what each command or sequence of commands accomplishes and what the purpose is.
DATA folder	● Goal: This folder contains all of the data for this project. ● Upload all datasets as a CSV file
OUTPUT folder	● Goal: This folder contains all of the output generated by your project, i.e. figures, tables, etc. ● Use informative names for your files ● Ensure each figure is labeled and summarized with a 1-2 sentence explanation
PRESENTATION folder	● Goal: This folder contains your final presentation deliverable as a pdf file ● The presentation can be made in any format i.e. powerpoint, poster, etc. ● The presentation must include: ○ Title ■ Name ■ Overview of topics ○ Motivation ■ Research question ■ Modeling approach ■ Background information ○ Analysis ■ Explanation of analysis method and why it was chosen ■ Tricky analysis decision ○ Results ■ At least one representation of data visualized i.e. figures, tables, etc. ■ Results from at least one statistical analysis test ■ Explanation of findings–how findings can be used by environmental researchers ○ Acknowledgements ■ All references should be listed at the end of the document ■ Use IEEE Documentation style (link)

Acknowledgements: Special thanks to my instructors Javier Rasero and Harsh Anand for their guidance throughout this semester. Thank you to my friends and group members Elizabeth Peterson and Mary Ellen Schuster for their hard work on each project we collaborated on together!