Introduction to the problem:

Understanding homelessness in San Diego requires looking beyond the policy and 311 call factors and examining the environmental pressures that shape the patterns of displacement. San Diego's climate is often characterized by warming temperatures, followed by recurring heatwaves, and seasonal fluctuations that are influenced by larger climate trends. This creates challenges that directly affect unsheltered populations. For instance, extreme heat events can worsen health risk, limit mobility, and strain public resources. In contrast, precipitation patterns and coastal climate may contribute to instability in housing markets and the availability of safe outdoor spaces.

This portion of the project investigates how climate data can help reveal trends in homelessness across San Diego, aiming to identify correlations between weather patterns and fluctuations in unsheltered populations. By integrating climate datasets and climate metrics with homelessness counts and spatial distribution data, the analysis seeks to uncover whether environmental stressors influence vulnerability or influence when and where homelessness is most visible.

The overall goal is to contribute to a more holistic understanding of homelessness that can support targeted, climate-responsive interventions and inform long-term planning for a region facing both rising housing pressures and a changing climate.

Datasets selected:
Custom GHCN from NCEI Climate Data Online request
ERA5-Land monthly averaged data from 1950 to present
Downtown San Diego Homeless Sleepers

The first dataset used in this project is the Global Historical Climatology Network-daily (GHCN-Daily for short). This dataset provides daily weather summaries dating back to 1939 that include a variety of variables, including minimum and maximum temperature, precipitation, and other meteorological conditions. Since this station has a long history of climate change records, it offers a reliable source of environmental context for understanding how weather patterns in San Diego fluctuate across different seasons and years. Although this dataset does not provide any insights into homelessness and trends tied to it, the data can be merged with homelessness datasets to see if there are correlations with the weather and environment alongside homelessness trends. For instance, seasonal temperatures can be analyzed to see if homelessness increases in hotter months, or if access to shelter is increased in months where precipitation is heavy. By integrating GHCN with social and homelessness data, this aims to meet the project's goals in the exploration of whether environmental conditions contribute to seasonal patterns or broader shifts in homelessness in San Diego.

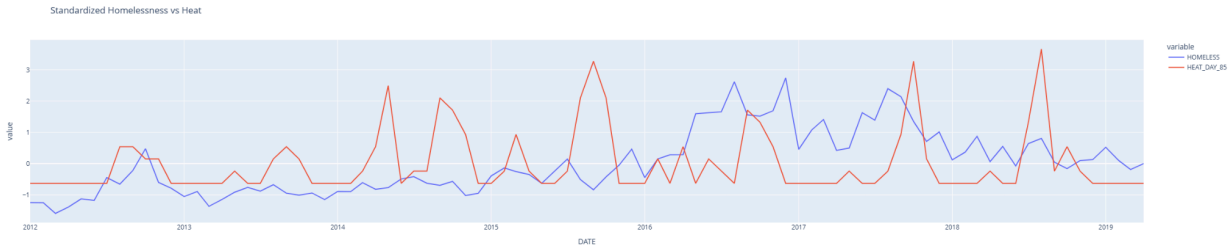
The second dataset used in this project is the ERA5-Land Monthly Averaged Reanalysis. This dataset is a globally gridded climate dataset produced by the Copernicus Climate Data Store. The ERA5 dataset provides a physics based reconstruction of land-surface climate conditions

from 1950 to the present day. This offers more monthly values for key environmental variables like the 2-meter air temperature, dewpoint temperatures, precipitation, soil moisture etc. Due to the fact that ERA5-Land integrates millions of observations into a climate model, it produces a coherent record, even in locations where time periods or weather station coverage are scattered. In the context of homelessness research, ERA5-Land provides the broader environmental backdrop needed to analyze how climate data may intensify homelessness in San Diego throughout the years. The dataset offers a monthly scale that allows researchers to examine patterns such as weather “hotter than average” months and how it corresponds to geographic shifts where individuals seek shade or cooler environments. Given the variety of variables the dataset has to offer, we can use variables like soil moisture, precipitation, and runoff to analyze how winter storm events can impact the safety and mobility of unhoused individuals living outdoors. In contrast with warmer weather, you can use dewpoints and temperature to calculate heat indexes to then see how heat-related health risks can impact unsheltered populations.

The third dataset used in this project comes from the San Diego Regional Data Library. This dataset is called “Downtown San Diego Homeless Sleepers” which provides roughly 5 years of monthly geographic positions and counts of individuals sleeping outdoors in Downtown San Diego. This dataset contains month-by-month counts from 2014 through 2019, with a few missing months noted by the documentation of the dataset. The dataset includes parameters and variables such as base counts, neighborhood-level totals, and census block geographies. For the analysis of homelessness trends in San Diego, this data provides the core social outcome measure needed to determine the correlation between climate and unsheltered populations. The monthly frequency allows researchers to detect short-term fluctuation, seasonal patterns, and neighborhood-level clustering that might be hidden in annual summaries. When combined with other datasets like GHCN-Daily and the ERA5-land, this dataset enables a richer understanding of how weather conditions, policy shifts, or seasonal cycles align with the movement in unsheltered homelessness, which is a core component of this project's objectives.

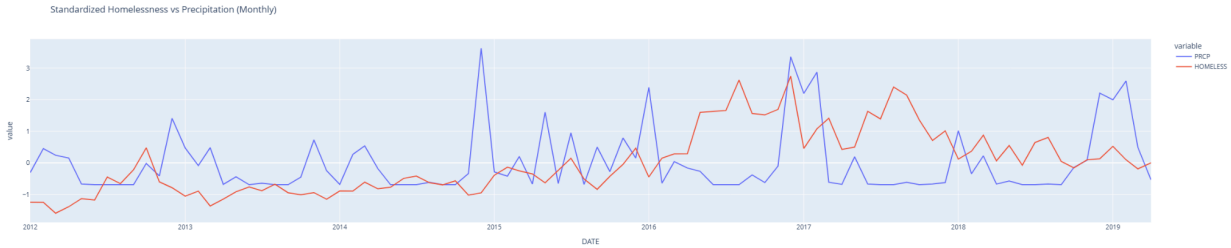
Generated Charts/Graphs/Explanation:

To explore the relationship between climate conditions and homelessness in San Diego, I first preprocessed the daily NOAA weather data by converting the raw station-level observations into monthly aggregates. Because of this simple conversion, I was able to clean the data and convert it to a time series format. I was able to calculate the key climate indicators like average monthly maximum temperature, precipitation totals, and number of heat days using resampling methods in pandas. The aggregated metrics were combined with the monthly homelessness counts from the Downtown San Diego dataset. This allowed me to create a unified analytical table covering both environmental and social variables across the same time span. To make the variables directly comparable despite their different scales, I standardized the homelessness counts, total heat days, and precipitation using z-score normalization. This step was crucial, as it allowed me to plot trends on the same axis without one variable visually dominating the other due to their differing magnitudes.



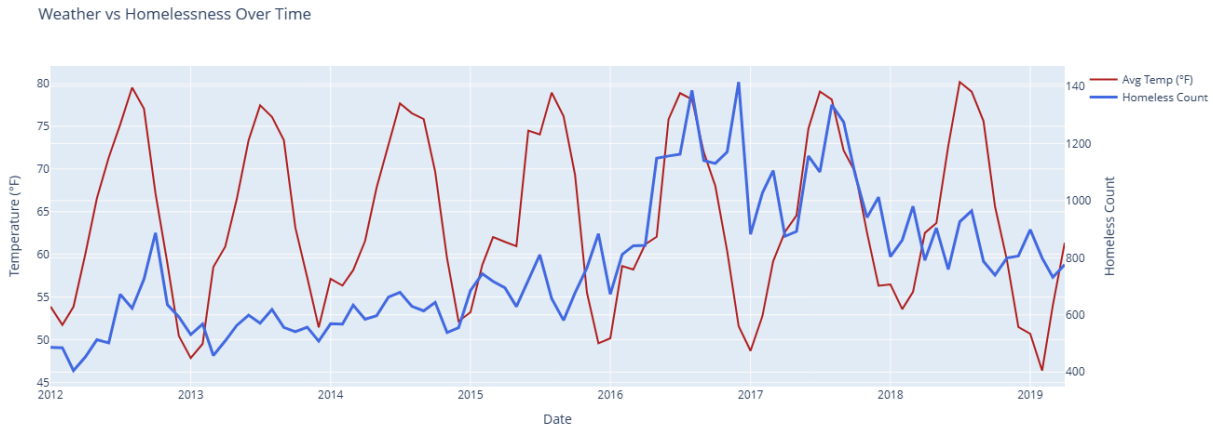
The figure above, “Standardized Homelessness vs. Heat,” visualizes how monthly fluctuations in homelessness align with periods of higher or lower heat exposure across several years. This figure suggests that there might be a loose association between hotter months and some noticeable shifts in unsheltered populations. Although the relationship between temperature and homelessness is not perfectly synchronized, the visualizations do reveal a recurring trend in the summertime temperatures alongside homelessness spikes.

To analyze whether rainfall patterns align with fluctuations in homelessness counts, I merged the monthly precipitation totals with the downtown homelessness dataset and standardized both variables using z-score normalization. After removing the missing values and creating a unified monthly timeline, I generated a line plot showing how each variable changes over time.



The resulting figure, “Standardized Homelessness vs Precipitation (Monthly)” reveals a bit more of an inverse relationship between precipitation and homelessness. This figure may suggest that rainfall may play more of a key role compared to heat, as individuals may seek more shelter during wet months. When the months are a bit warmer or drier, unsheltered populations may find it a bit more feasible to stay outdoors. While these two figures above do not explain the causation of homelessness, they do offer a highly noteworthy seasonal dynamic between rainfall and unsheltered populations.

To now see how I can extend my observations with homeless populations, I merged the monthly ERA5-Land air temperature data with the Downtown San Diego Monthly homeless totals. The temperature values were extracted from the netCDF reanalysis datasets. I then converted them from Kelvin to Fahrenheit and aggregated them to monthly averages. The homelessness data was similarly normalized to monthly timestamps as well to ensure that both datasets aligned on the same temporal scale. After merging, the other figures I implemented a dual-axis time series plot with temperature displayed on the y-axis and homelessness on the secondary y-axis.



The resulting figure shows a clear and recurring pattern that we have seen with the other two figures. However, this does show how homelessness increases during warmer months and declines during cooler months. The temperature follows a cycle in San Diego where it peaks during late summer and dips during the winter. Going back to one of my original points, is that unsheltered populations may find the warmer weather more feasible to sleep outside, whereas in contrast, they may seek to find shelter during wet weather or heavy precipitation.

Constraints and limitations:
Although the figures that have been produced in this project do give us key information on how unsheltered populations trend in certain climates. Using similar datasets may give us key information on where they migrate, more maps and graphs to be generated, and so forth. The big problem behind that is due to the fact that a lot of the time, climate data is extremely unstructured. The process of reindexing climate data to a database becomes an issue because while working with a netCDF file that is a total of 4GB can grow to something that is about 40GB. The tradeoff with that is that, despite how big the dataset may be, databases are usually lightweight and are fast when it comes to processing. In contrast, when you are with the raw netCDF file, it takes up a lot of computing power to generate said graphs and maps. The last issue I ran into with climate data is that a lot of it is poorly documented, so finding the right datasets for this project was extremely difficult. With the homeless datasets I have gathered, there was a lack of information in certain months; I had to filter out when cleaning, which might have affected the results in some way that could be misleading. Like the climate datasets, I had a difficult time finding homelessness datasets I could use alongside the climate ones, solely because a lot of them did not have information from 2000 - 2025.

Conclusion:
To answer the question on how climate affects homelessness, given the information, first, it must be known that climate change DOES NOT cause homelessness. The accurate conclusion is that during warmer temperatures, homeless individuals tend to be outside more, just because they find the warmer weather to be more feasible for sleeping outside. During cooler weather, winter time, or heavy rainfall, the homeless population tends to decrease. This does not suggest

that the homelessness problem is being fixed, but rather, that there is more access to shelter during these times. This idea is suggested solely because when the weather gets warmer again, the homeless population tends to spike up a bit more. This has been the ongoing cycle when it comes to analyzing the trends and patterns of homelessness in correlation to the weather in San Diego.