Capital One - Rain Check

Overview
Our sponsors, Capital One, strives to improve banking for their customers. One way Capital One is trying to better their customer experience by protecting their customers’ from fraudulent purchases by improve their fraudulent detection system. Our initial problem is: can Capital One use weather data to provide additional insight into how their customers behave and spend to help better their fraud detection scoring model.

Objectives
Our team’s object is to improve Capital One’s fraud scoring model, our design will be used to determine whether spending habits change for certain types of events depending on the customer's geographic regions and weather types. To find a relationship between customer spending habits and weather, we gathered purchase data and weather data from Walmart to see if there was some type of correlation. Then we will build a model to find a trend in the customer purchase data during temperature changes as well as severe storms.

Approach
1. We acquired data of Walmart sales from 99 departments spanning 45 stores off of kaggle.com
2. The holiday sales data were removed as holidays often have spending that does not follow seasonal trends
3. The data was split into 4 seasons based on months
4. In order to account for the transition between seasons, the first and last weeks of each season were removed
5. Using scipy.stats.mean and scipy.stats.nanstd we calculated the means and standard deviations for each department in each store.
6. We then calculated the z-score using mean and standard deviation for both the temperature and sales for each data point. The z-score allows us to analyze difference from the mean. This allows us to analyze how unseasonably high/low temperatures affect weekly sales.
7. We then cleaned the data of sets where there were no data points or only one data point.
8. Using scipy.stats.linregress, we found linear functions to model the data that was presented.
9. We then created graphs with the linear regression line and the data points using matplotlib.

Outcomes
- The data suggests that there is indeed a relationship between changing temperatures and sales.
- Using our data model other metrics such as rainfall or wind speed could be compared as well.
- We believe that Capital One now has a reasonable incentive to further investigate this problem using real customer sales data.