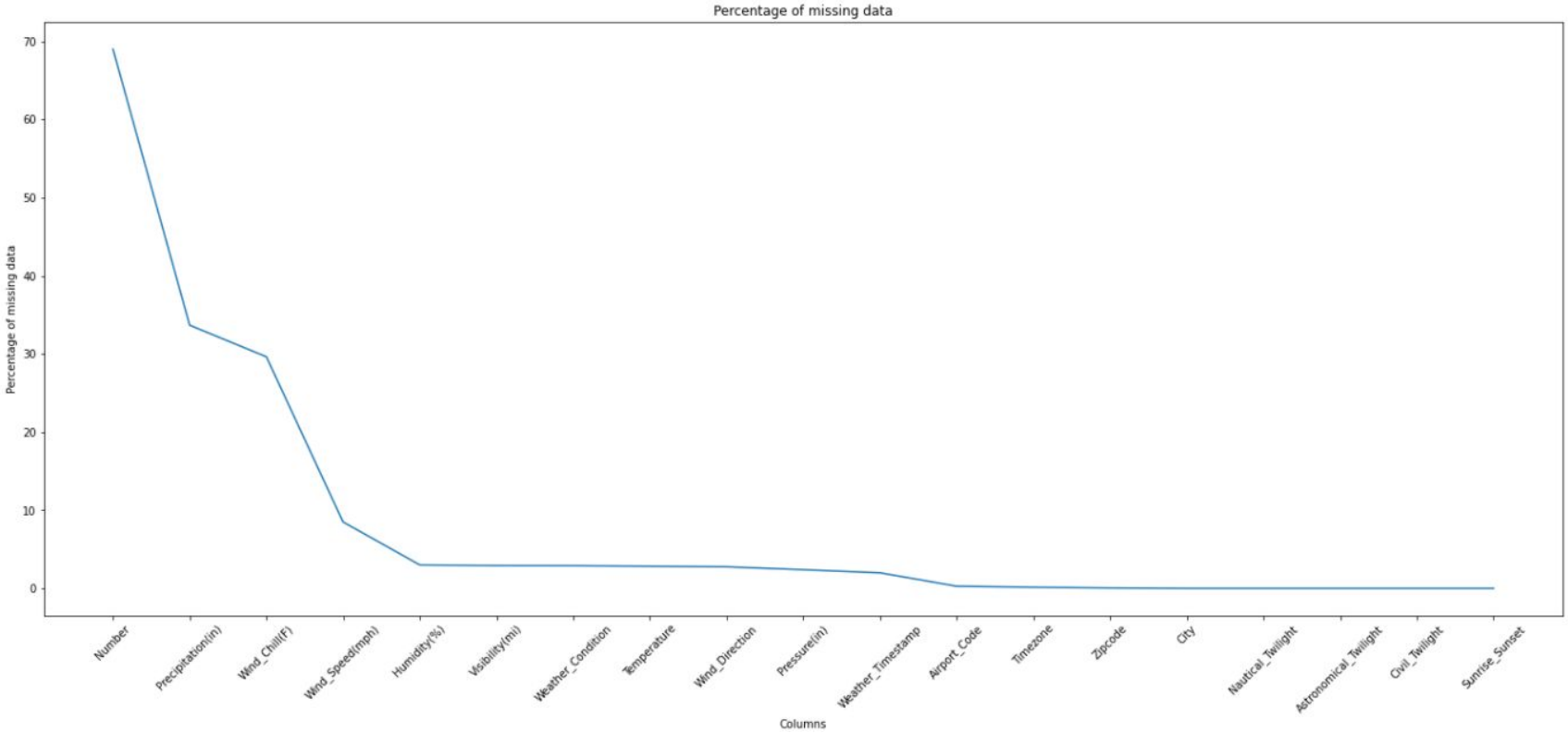
USA Accidents - EDA

Analysing the data to determine what factors affect the number of accidents in USA

About the dataset

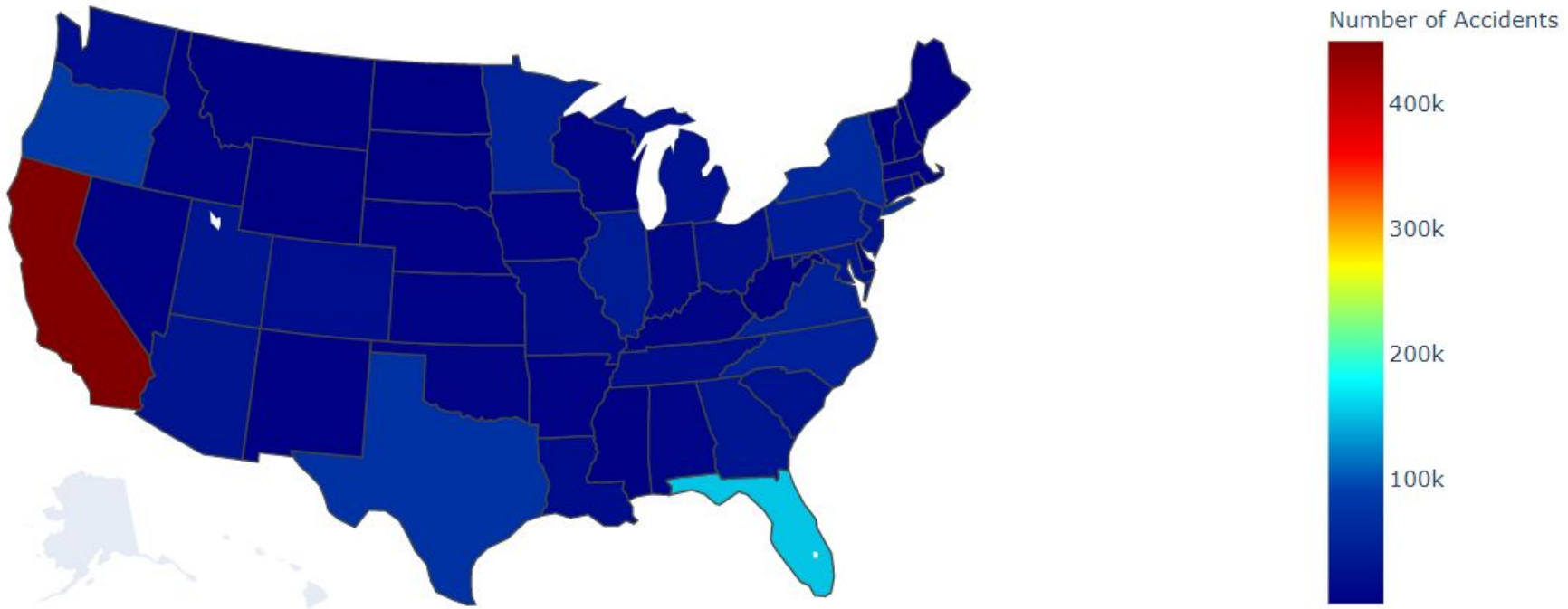
This data is obtained from Kaggle.This is a countrywide car accident dataset, which covers 49 states of the USA. The accident data are collected from February 2016 to Dec 2020, using multiple APIs that provide streaming traffic incident (or event) data. These APIs broadcast traffic data captured by a variety of entities, such as the US and state departments of transportation, law enforcement agencies, traffic cameras, and traffic sensors within the road-networks. Currently, there are about 1.5 million accident records in this dataset.

Data Preparation and Cleaning

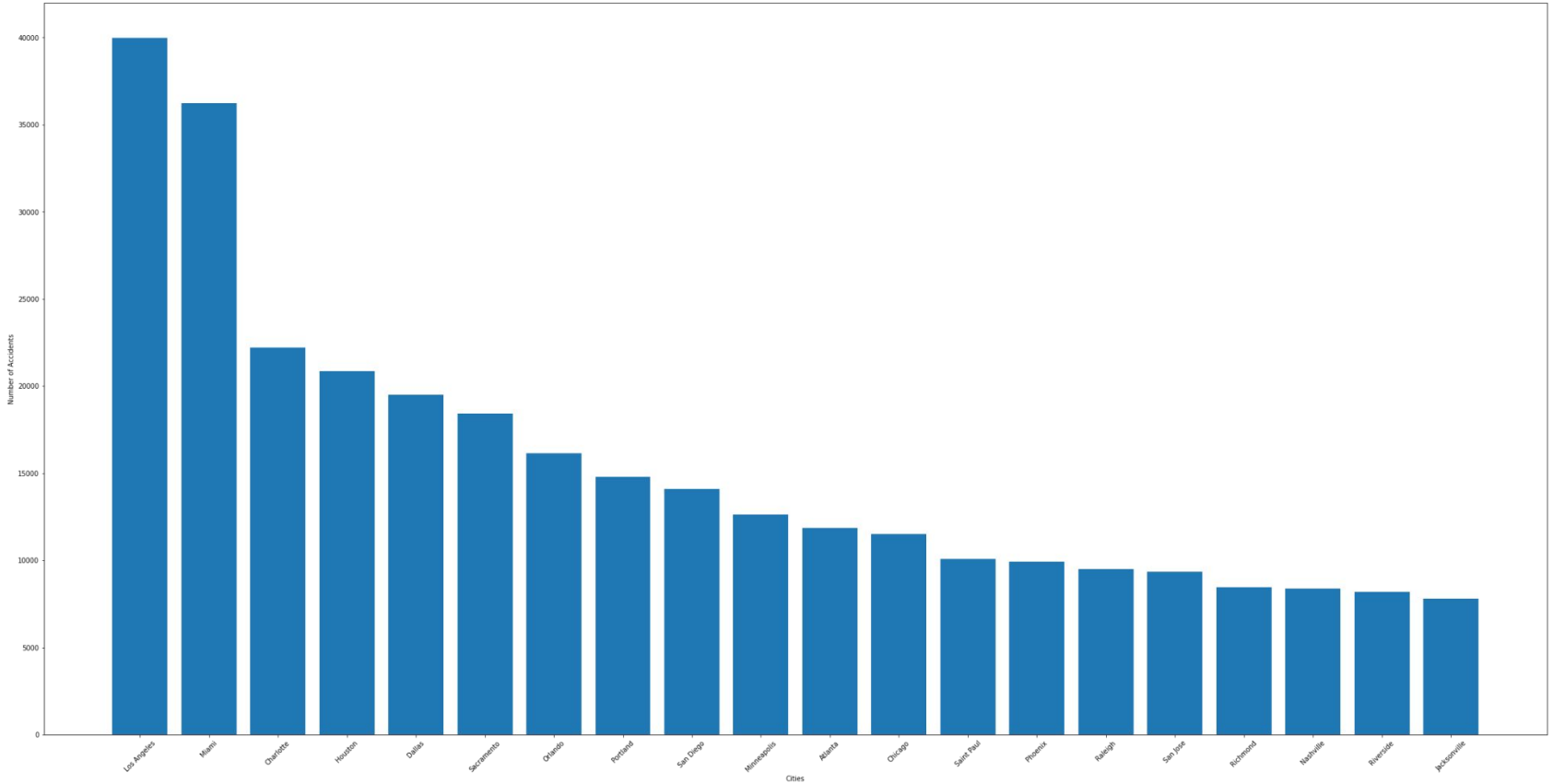


The number of missing values in Precipitation, Wind_Chill & Wind_Speed is relatively high. Hence, these columns can be omitted.

Most Dangerous States

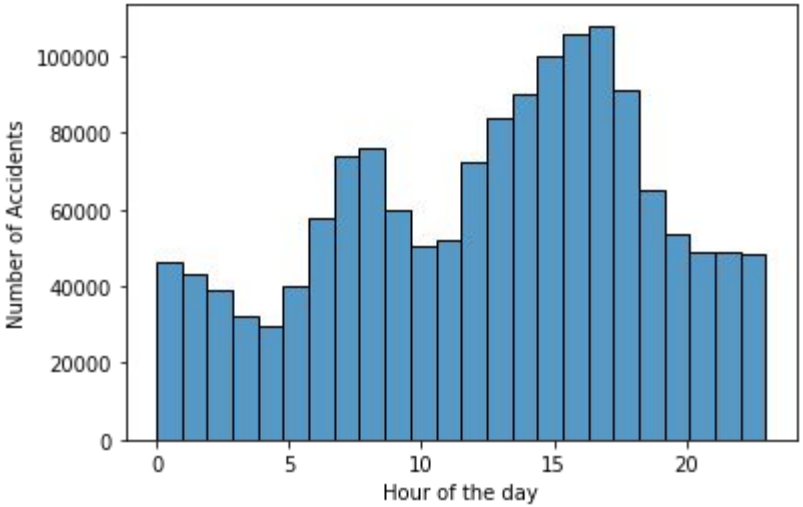


Most Dangerous Cities



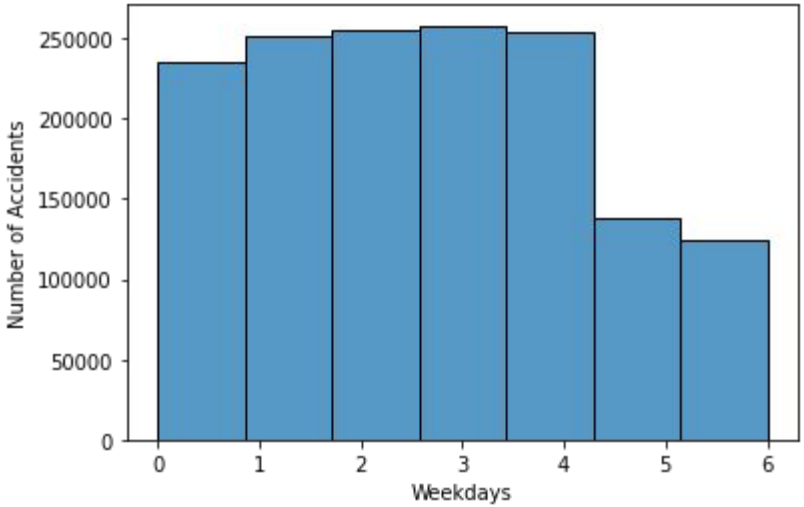
Most prone time of accidents

It seems like most accidents occur during 7AM to 10AM & during 1PM to 8PM. The most logical explanation for hike in accidents is because of traffic rush due to working people. Let's check if the trends are same on weekdays and weekends.

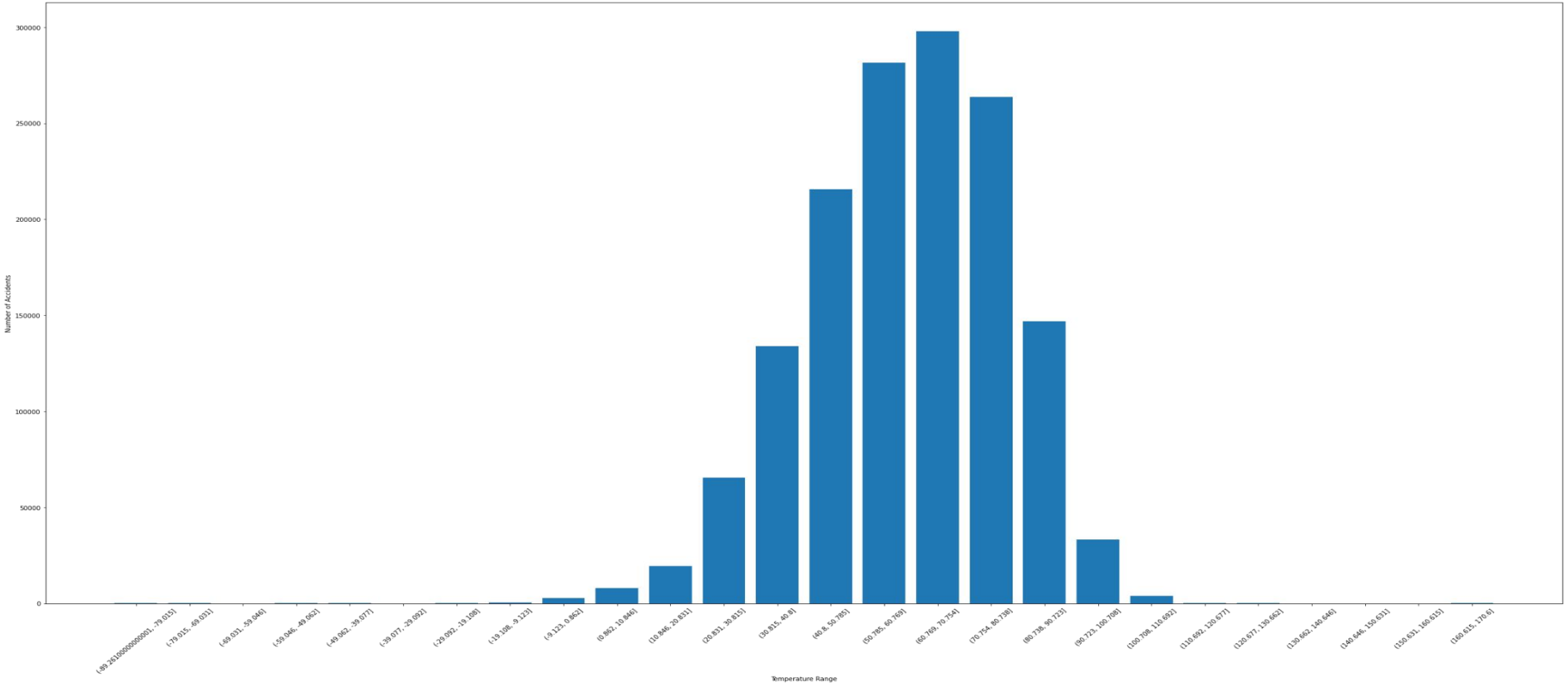


Most prone day of accidents

As we can see here, most number of accidents occur on weekdays which again supports our theory regarding working class.



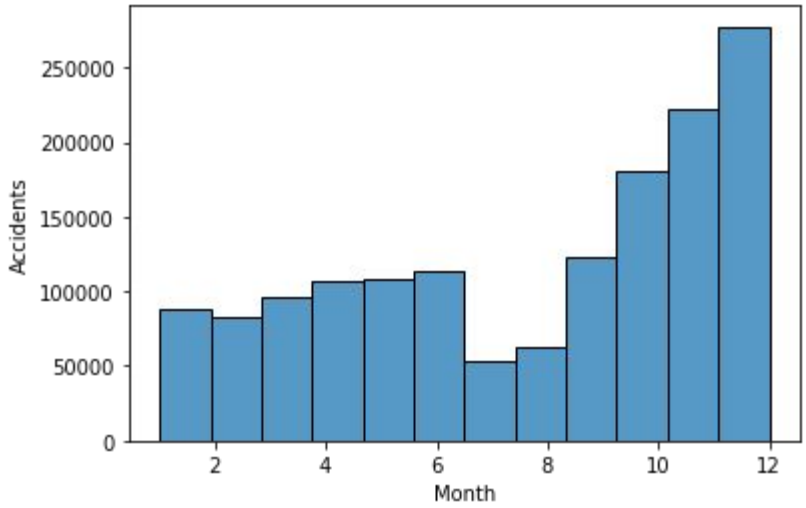
How does weather & temperature affect the accidents?



Most prone month of accidents

From the graph above we see that most of the accidents occur within thee temperature range of 40°F to 90°F which is a significantly cool weather.Let's see we can deduce a more meaningful insight from the variation per month.

This shows that most of the accidents occur from October to December. This verifies our theory that winter season plays a major role in accidents. Maybe due to harsh weather conditions and low visibility,accidents are more.



Future Analysis

1.After combining another dataframe for each city's population,we can calculate the number of accidents per head to determine where the density of accidents is highest.

2.The data about the road conditions can be analysed to determine where the road requires management.