



US CAR ACCIDENT ANALYSIS

SUBMITTED BY BRUTUS LO



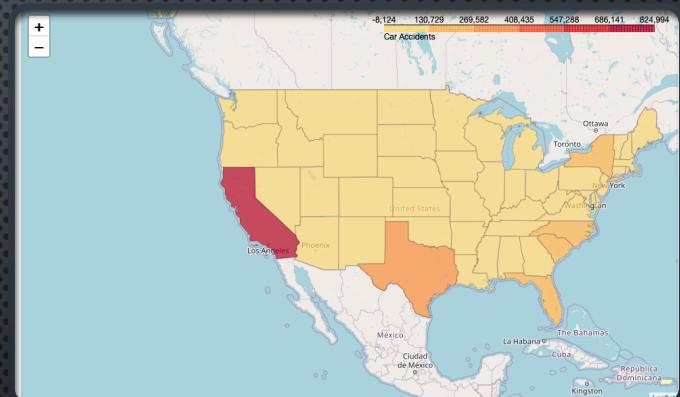
INTRODUCTION

PERSPECTIVES OF AN AUTO INSURANCE COMPANY

- FREQUENCY AND SEVERITY OF CAR ACCIDENTS ARE IMPORTANT TO KNOW.
- WHAT ARE POSSIBLE FACTORS AFFECTING SEVERITY OF A CAR ACCIDENT?
 - WEATHER PATTERNS
 - WEATHER FACTORS
 - PRESENCE OF POINT OF INTEREST
- THIS REPORT ONLY FOCUS ON WEATHER FACTORS

DATASET

- DATASET COVERS 49 CONTIGUOUS STATES OF THE USA FROM FEB 2016 – JUNE 2020
- TOTALING 3.5 MILLION RECORDS
- CREDITS/ACKNOWLEDGEMENTS:
- MOOSAVI, SOBAN, MOHAMMAD HOSSEIN SAMAVATIAN, SRINIVASAN PARTHASARATHY, AND RAJIV RAMNATH. "A COUNTRYWIDE TRAFFIC ACCIDENT DATASET.", 2019.
- MOOSAVI, SOBAN, MOHAMMAD HOSSEIN SAMAVATIAN, SRINIVASAN PARTHASARATHY, RADU TEODORESCU, AND RAJIV RAMNATH. "ACCIDENT RISK PREDICTION BASED ON HETEROGENEOUS SPARSE DATA: NEW DATASET AND INSIGHTS." IN PROCEEDINGS OF THE 27TH ACM SIGSPATIAL INTERNATIONAL CONFERENCE ON ADVANCES IN GEOGRAPHIC INFORMATION SYSTEMS, ACM, 2019.

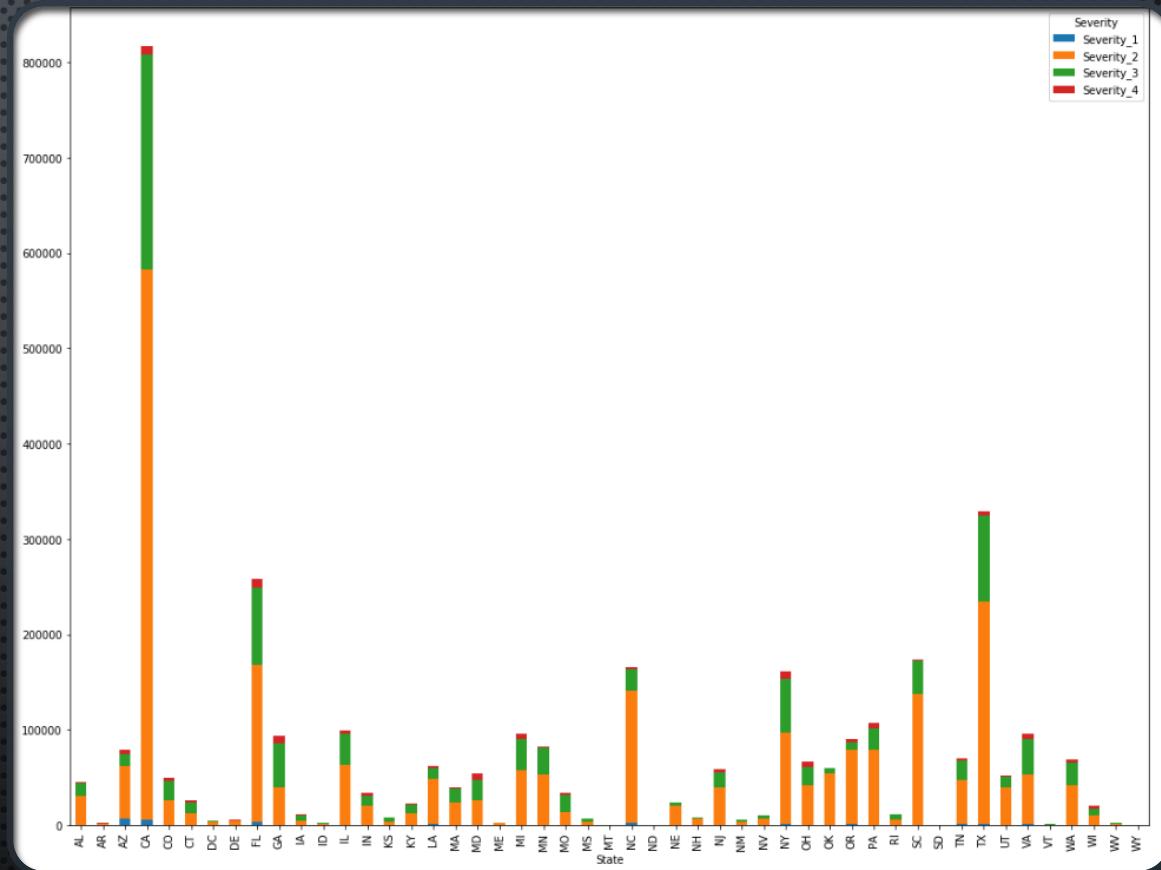


DATA CLEANUP

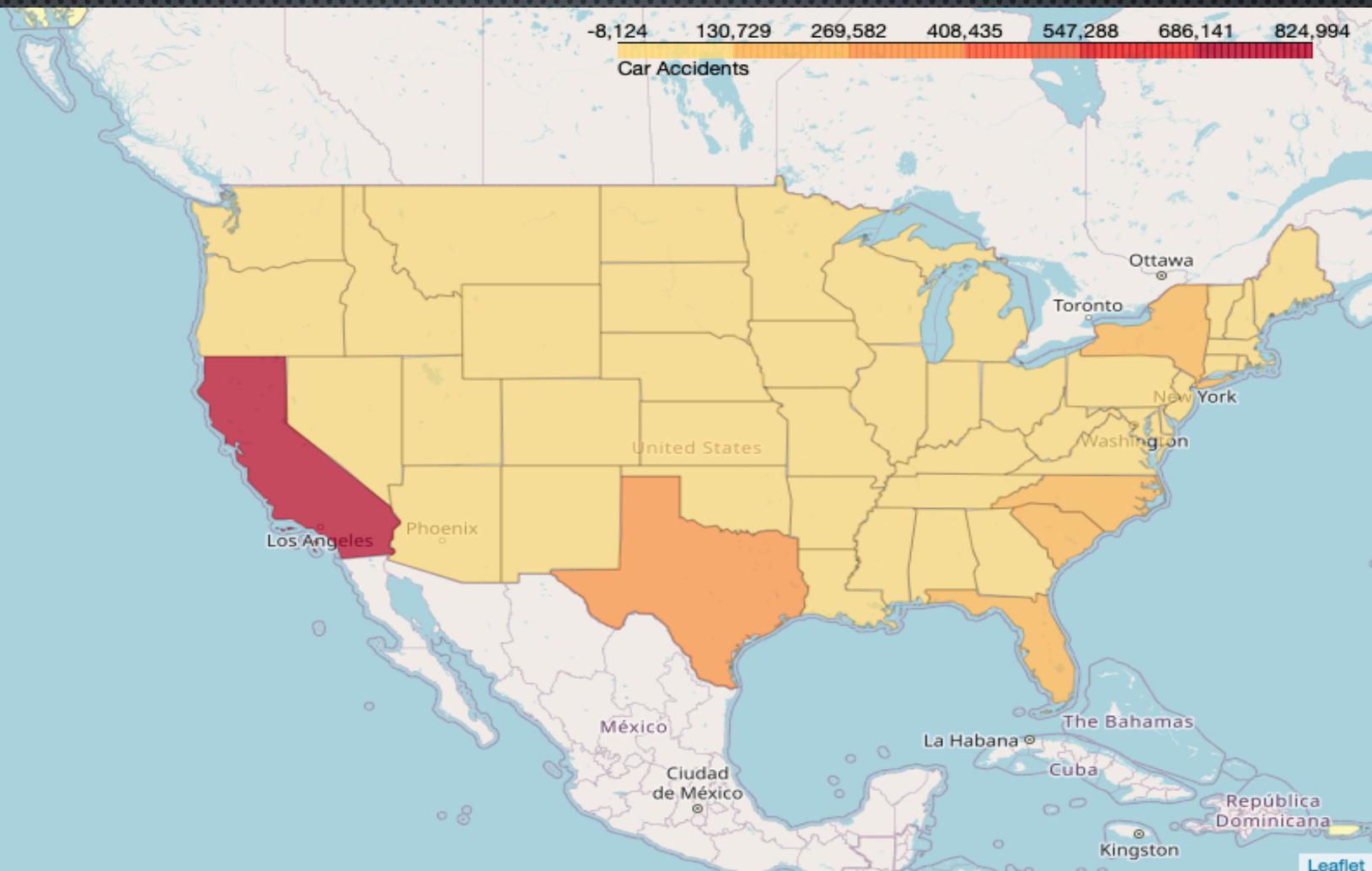
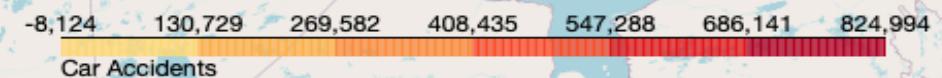
- ORIGINAL DATASET HAS 49 ATTRIBUTES
- FOCUSING ONLY ANALYZING WEATHER CONDITION AND PATTERN
- ELIMINATE UNNEEDED DATA USING .DROP
- FINDING MISSING DATA THEN USING .REPLACE OF NAN BY THE AVERAGE VALUES

OVERVIEW: ACCIDENT SEVERITY BY STATE

- CALIFORNIA HAS THE HIGHEST # OF ACCIDENTS AND HAS MORE THAN DOUBLE THE FREQUENCY OF ACCIDENTS THAN THE NEXT STATE, TEXAS

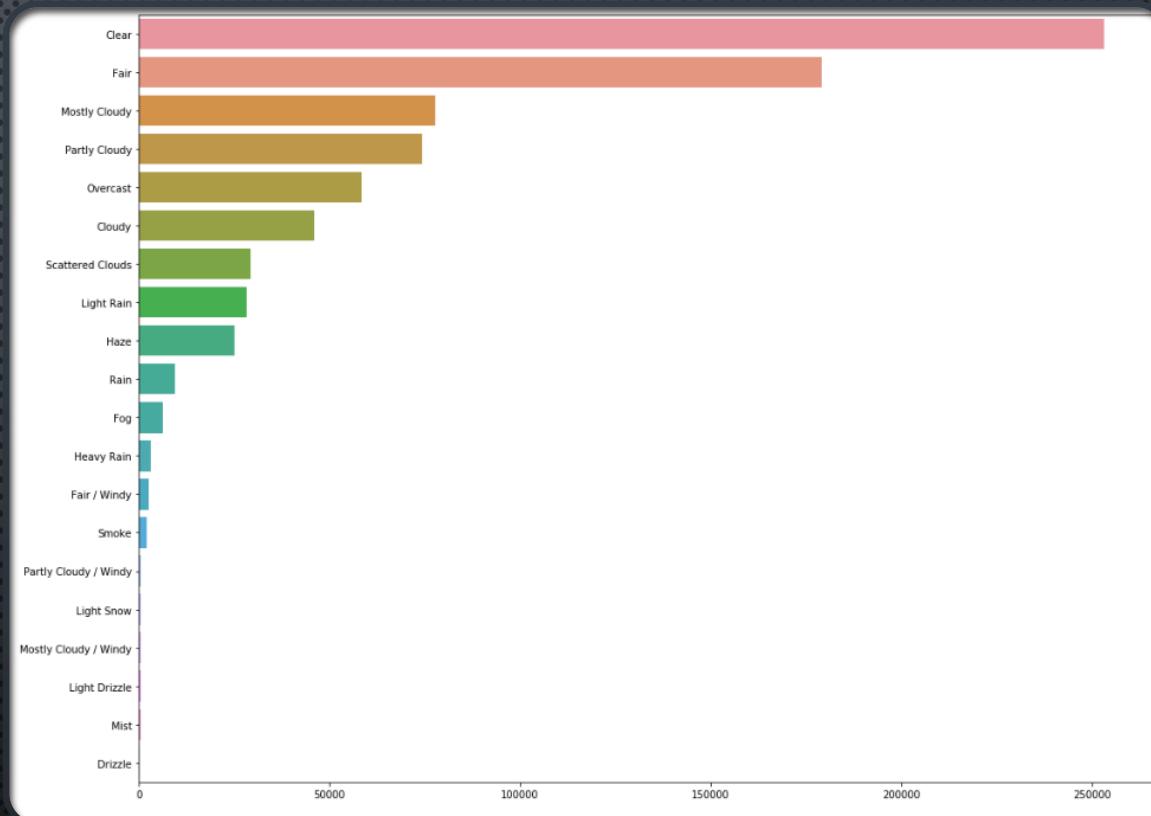


ACCIDENT DISTRIBUTION



ACCIDENT FREQUENCY IN VARIOUS WEATHER CONDITIONS: CA DATA ONLY

- SURPRISINGLY, MOST ACCIDENTS OCCURRED IN CLEAR WEATHER CONDITION.
- VERY SMALL NUMBER OF ACCIDENTS OCCURRED IN SNOWING, FOGGY OR WHAT WE WOULD CALL “POOR” CONDITIONS.



WEATHER FACTORS

- WINDSPEED
- TEMPERATURE
- WIND CHILL
- HUMIDITY
- VISIBILITY
- PRESSURE

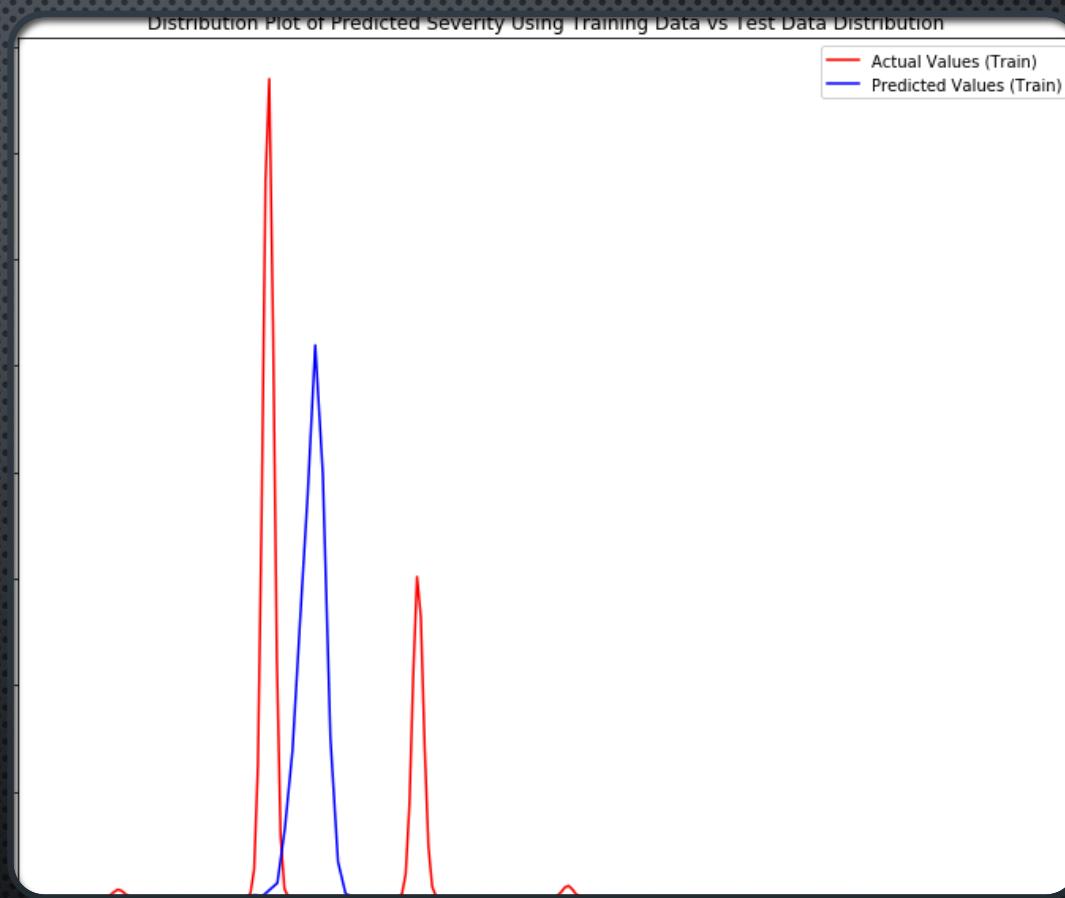


LINEAR REGRESSION

WEATHER FACTORS	Pearson Coefficient; p-value
Windspeed	0.041, 4.731e-305
Temperature	0.024, 5.989e-104
Wind Chill	-0.108, 0.0
Humidity	0.030, 4.284e-157
Visibility	-0.005, 4.008e-06
Pressure	0.095, 0.0

MULTIPLE REGRESSION

- TEST SIZE: 20%
- ACTUAL AND PREDICTED VALUES ARE FAR OFF



RIDGE REGRESSION & GRID SEARCH

- RIDGE REGRESSION RESULTS:
 - PREDICTED: [2.15239507 2.47268982 2.34856686
2.23223706] TEST SET : [2 2 3 2]
 - BEST RR SCORE USING VARIOUS ALPHA: .030

CONCLUSIONS & FURTHER STUDIES

- WEATHER FACTORS DO NOT GIVE GOOD PREDICTION OF SEVERITY OF A CAR ACCIDENT
- FURTHER STUDIES:
 - PRESENCE OF POINT OF INTERESTS
 - TIME OF THE DAY