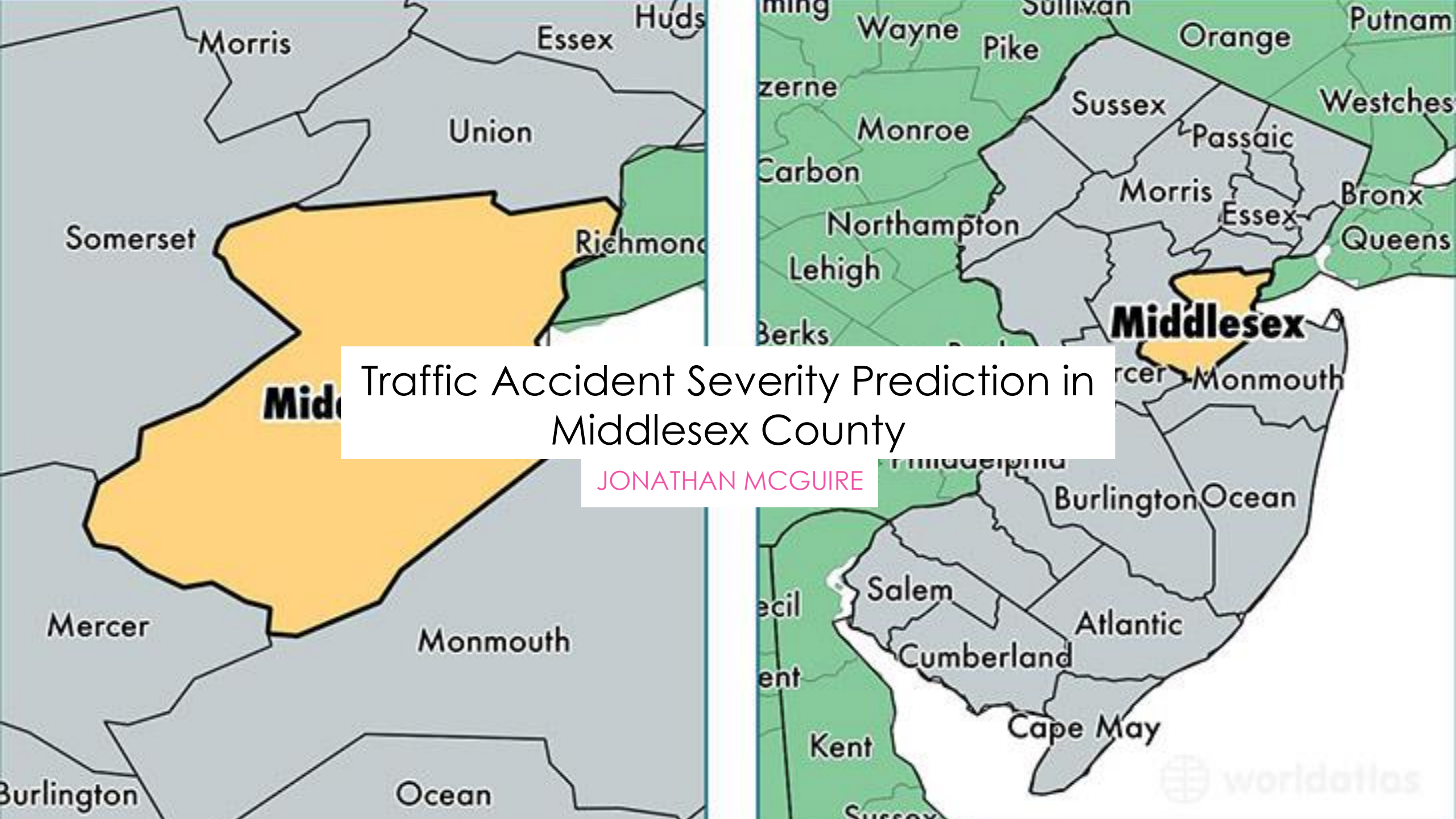


Traffic Accident Severity Prediction in Middlesex County

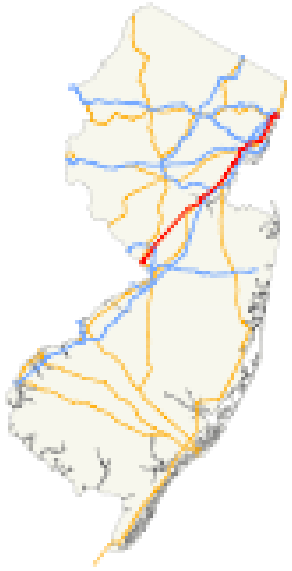
JONATHAN MCGUIRE

Introduction

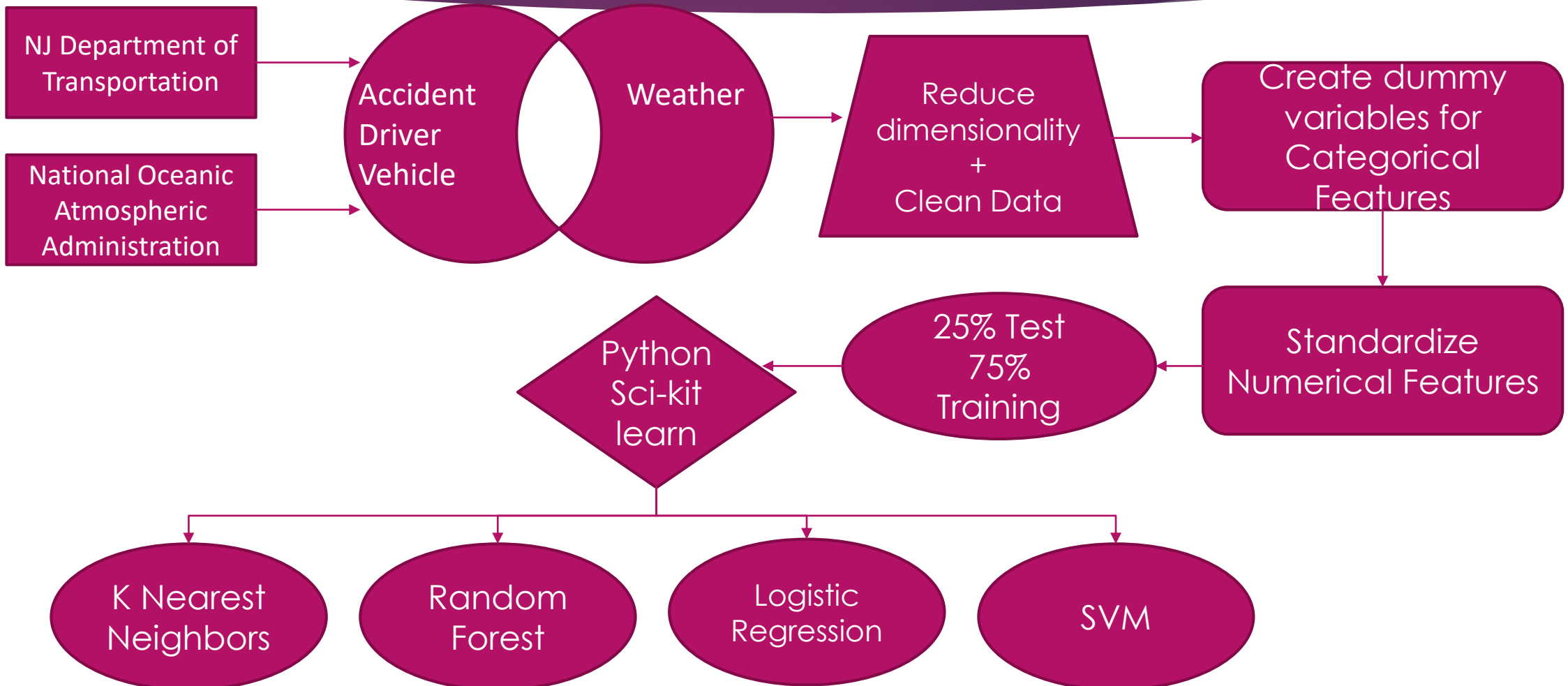
- “A total of 35,092 people died in motor vehicle crashes in 2015. The U.S. Department of Transportation's most recent estimate of the annual economic cost of crashes was $242 billion dollars.”-IIHS
- “In 2013, the US crash death rate was more than twice the average of other high-income countries.”-CDC
- “One in 3 crash deaths in the US involved drunk driving, and almost 1 in 3 involved speeding.”-CDC

Objectives

- Find a model that best aids future prediction of accident severity
- Examine historical data to find influence of particular attributes
- Influence future policies on preventing motor vehicle accidents



Process Flow

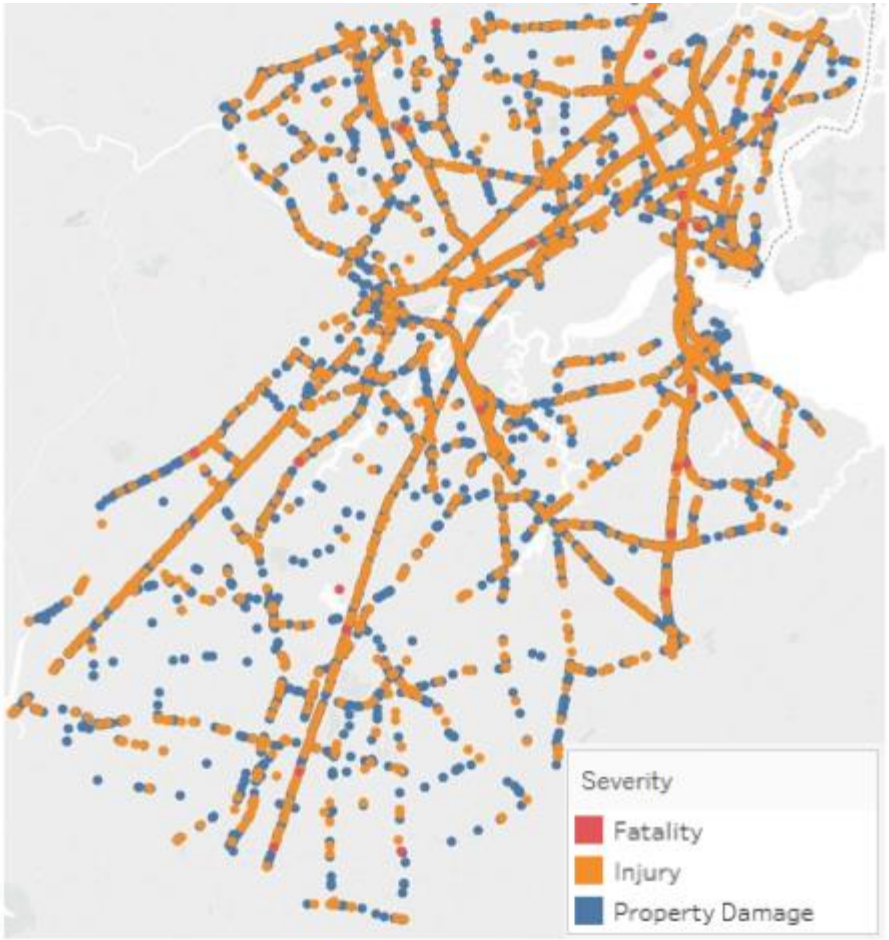


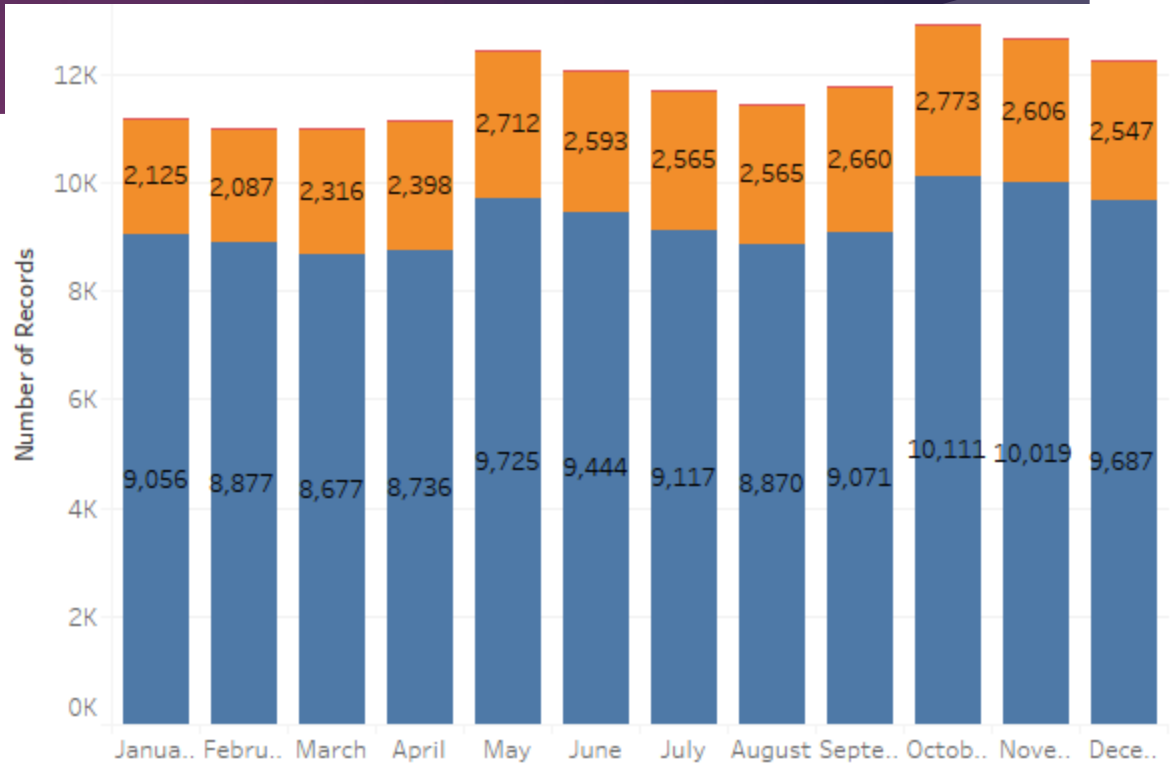
Data Overview

- 140,000~ Instances
- 100~ attributes -> 14 attributes

Month	alcohol involved	total vehicles involved	road system	light condition	road divided by	posted speed	cell phone in use flag	Precipitation	Snowfall	Temp Average	Age	year of vehicle	severity
August	N	2	3	1	1	55	N	0	0	73			1
July	N	2	3	1	1	15	N	0.04	0	82.5			1
July	N	3	3	1	1	55	N	0		80			1
January	N	2	2	1	5	55	N	0.3	0	37.5	12	2004	1
October	N	2	7	1	5	0	N	0.01	0	56	13		1
Novembe	N	2	7	1	5	25	N	0.84	4.8	35.5	13	2003	2
April	Y	1	3	5	1	65	N	0.33	0	66	13	2000	1
May	N	1	9	6	5	0	N	0	0	57.5	14	2007	1
March	N	1	9	1	5		N	0.09	0	32.5	14	1995	1
June	N	2	9	7	5		N	0.04	0	70.5	14	2008	1

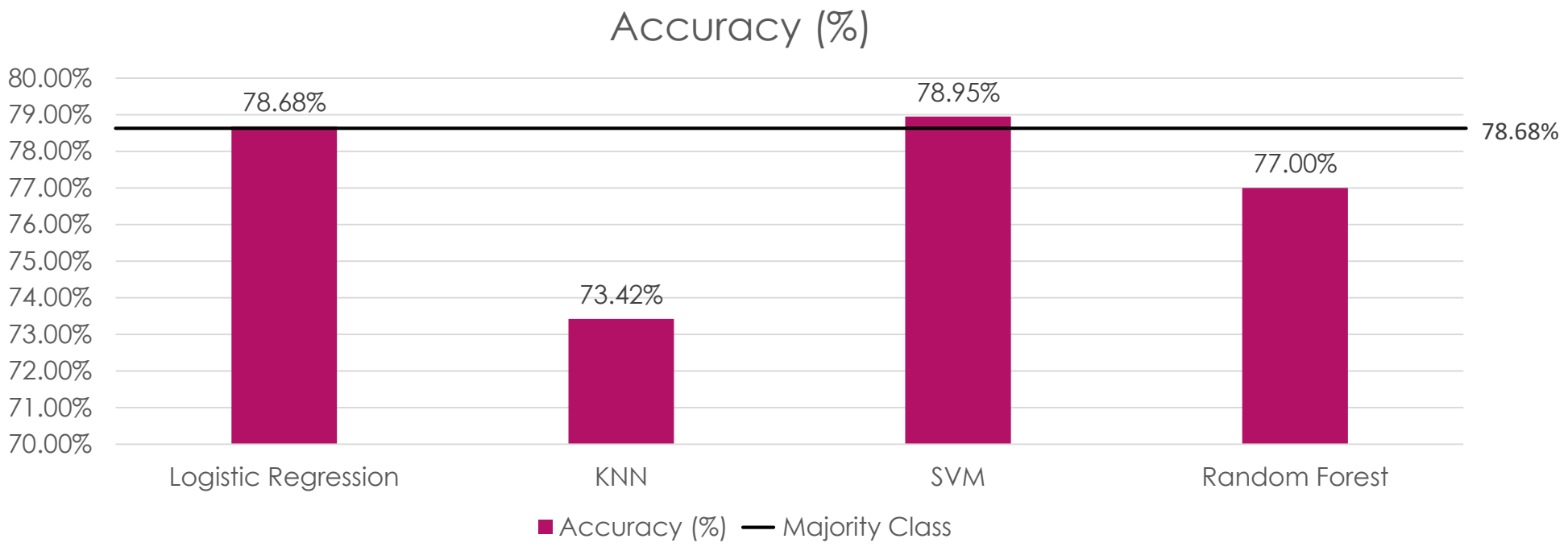
Data Overview





Severity		
Fatality	234	0.17%
Injury	29,990	21.16%
Property	111,510	78.68%

Algorithms Comparison



Log Odds

Variable	Property	Injury
Month	1.0014	1.0021
alcohol involved	0.1669	0.4087
total vehicles involved	2.2815	2.8116
road system	0.9893	0.8981
light condition	0.8687	0.8927
road divided by	1.0206	1.0141
posted speed	0.9633	0.9664
cell phone in use flag	0.318	0.4453
Precipitation	1.0839	1.0584
Snowfall	0.9964	0.9578
Temp Average	0.9956	0.9993
Age	0.9751	0.977
year of vehicle	1.0591	1.0358

Project Task Log

Date	Time start	Time End		Hours	Description
29-Oct	2:30 PM	4:30 PM	2:00	2	Dataset Search
2-Nov	10:00 AM	12:50 PM	2:50	2.5	Dataset Search
4-Nov	10:00 AM	11:00 AM	1:00	1	Preliminary Outline
5-Nov	2:00 PM	5:00 PM	3:00	3	Introduction+Obtaining+Cleaning Data Sections
7-Nov	10:00 AM	2:00 PM	4:00	4	Literature Review
7-Nov	6:00 PM	6:30 PM	0:30	0.5	Citations
8-Nov	10:00 AM	2:00 PM	4:00	4	Evaluation+Approach
8-Nov	6:00 PM	8:00 PM	2:00	2	Power Point
8-Nov	9:00 PM	11:30 PM	2:30	2.5	Revision
1-Dec	9:00 PM	12:00AM	3:00	3	Data cleaning
2-Dec	2:00 PM	3:00 PM	1:00	1	Data cleaning
2-Dec	3:30 PM	6:00 PM	2:30	2.5	Data cleaning
2-Dec	7:00 PM	9:00 PM	2:00	2	Data cleaning
2-Dec	10:30 PM	11:30 PM	1:00	1	Poster Outline
3-Dec	10:30 PM	12:30 AM	2:00	2	Data cleaning
4-Dec	1:00 AM	2:30 AM	1:30	1.5	Data cleaning
5-Dec	1:00 PM	5:00 PM	4:00	4.00	Poster Work
6-Dec	10:00 PM	2:00 AM	4:00	4.00	Python
8-Dec	6:00 PM	8:00 PM	2:00	2.00	Python
9-Dec	1:00 PM	6:00 PM	5:00	5.00	Python
10-Dec	1:00 PM	3:00 PM	2:00	2.00	Python
11-Dec	2:00 PM	4:00 PM	2:00	2.00	Python
13-Dec	1:00 PM	4:00 PM	3:00	3.00	Powerpoint