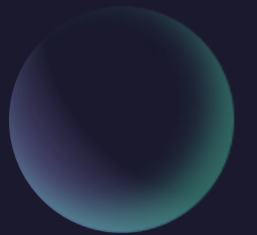


Capstone Project - Car accident severity

IBM | Data Science Professional Certificate



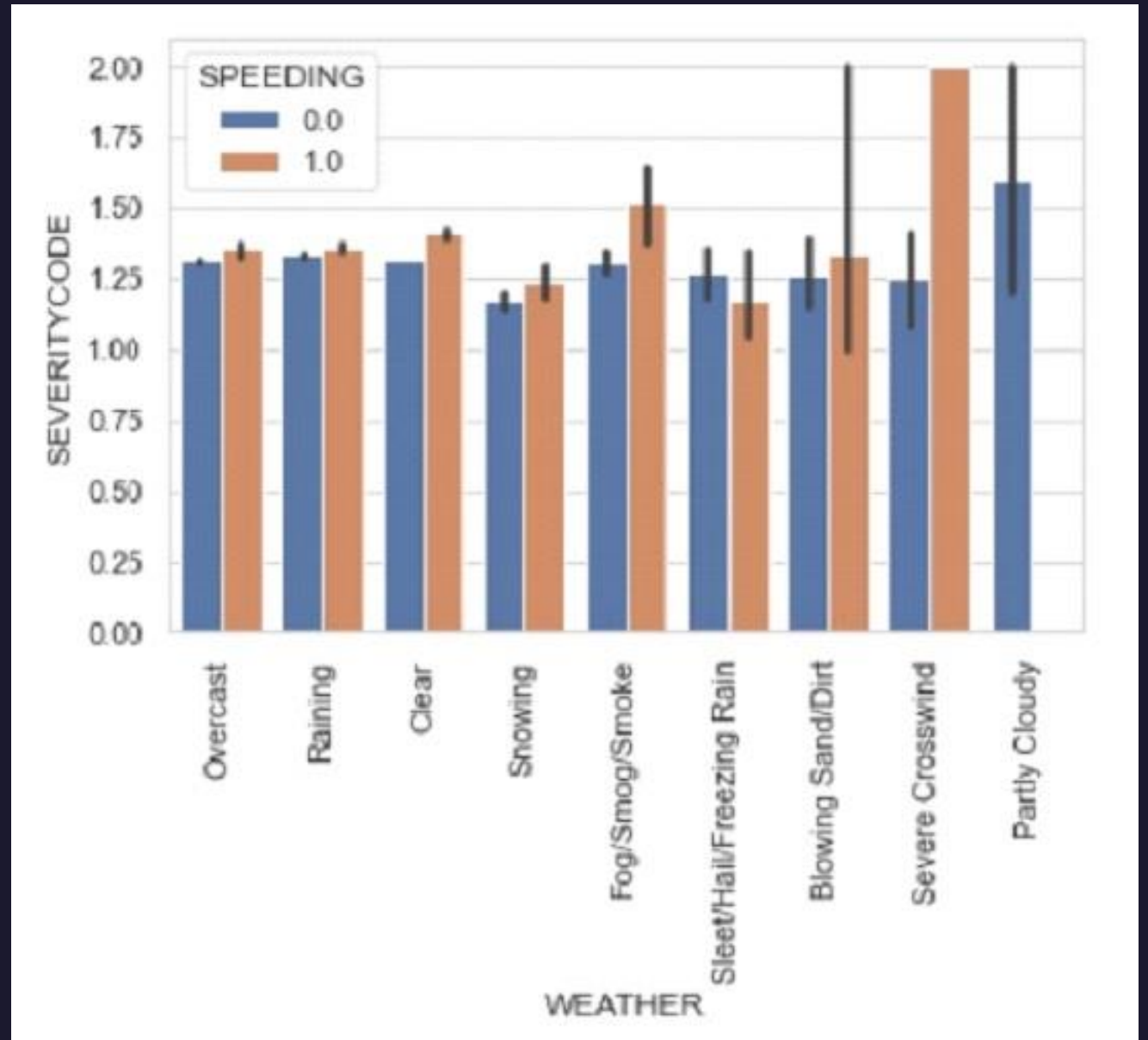
Knowing more reasons why accidents happen can save lives.

- ❖ Most car accidents are caused by driver speed. But apart from that, there are many different factors that cause accidents.
- ❖ Finding these factors accelerates taking measures to reduce the rate of accidents.

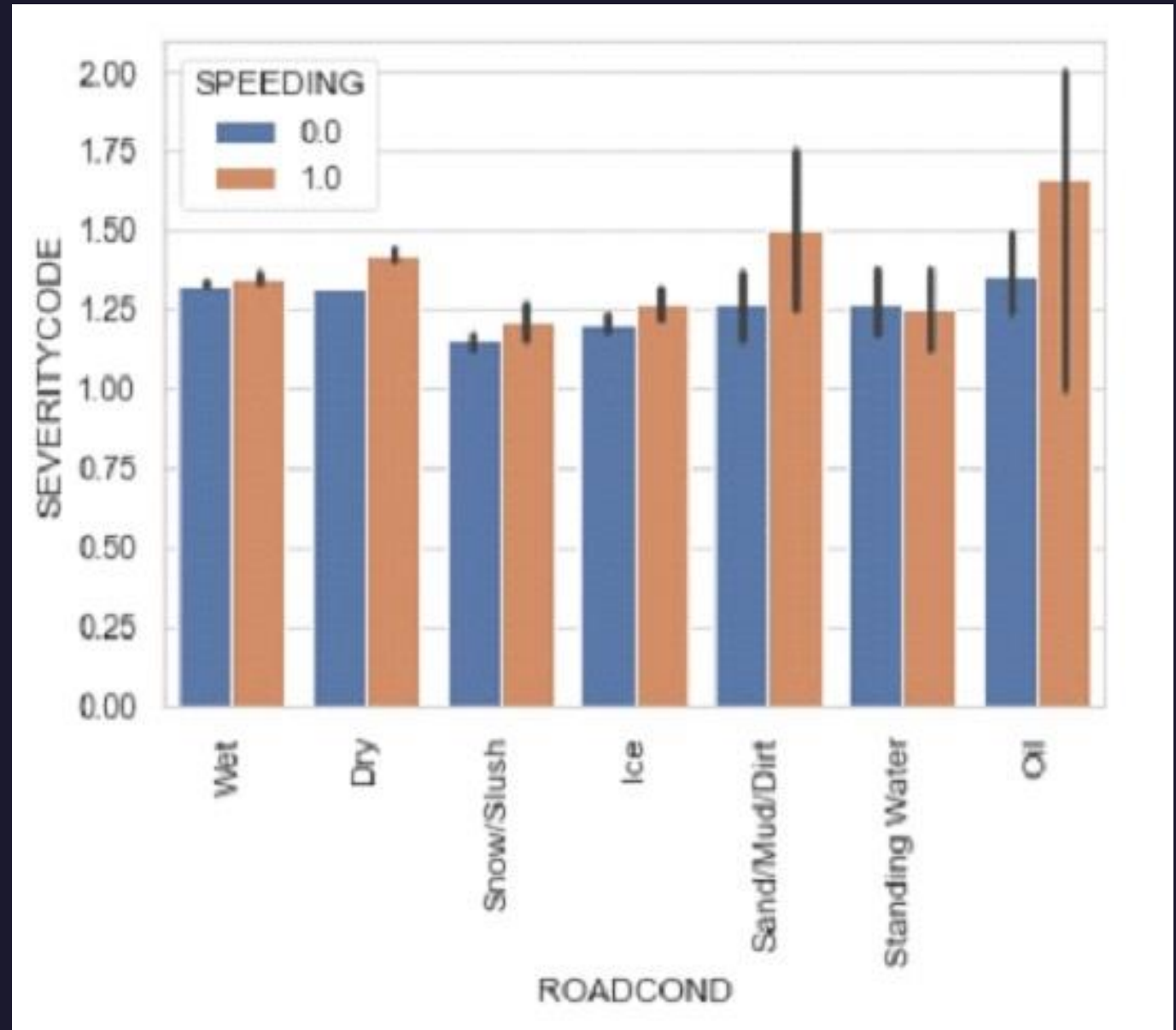


- ❖ The data has been provided by the Seattle Police Department since 2004 and consists of 37 independent variables and 194.673 lines.
- ❖ In order to find out which factors affected the injuries in accidents, firstly the column "SEVERITYCODE" which is the dependent variable and defines the severity of the collision and "COLLISIONTYPE" which defines the type of collision that will change it, "WEATHER" which defines the weather at the time of the collision, "ROADCOND" which describes the road condition and light conditions. I chose to use in the dataset the columns "LIGHTCOND", "INATTENTIONIND", which describes whether the driver is distracted, and "SPEEDING", which describes the speed of the driver.
- ❖ "SEVERITYCODE" contains numbers from 0 to 4 corresponding to different severity levels of the accident as:
 - 0: Little to no injured.
 - 1: Very Low injured.
 - 2: Low injured.
 - 3: Mild injured.
 - 4: High injured.

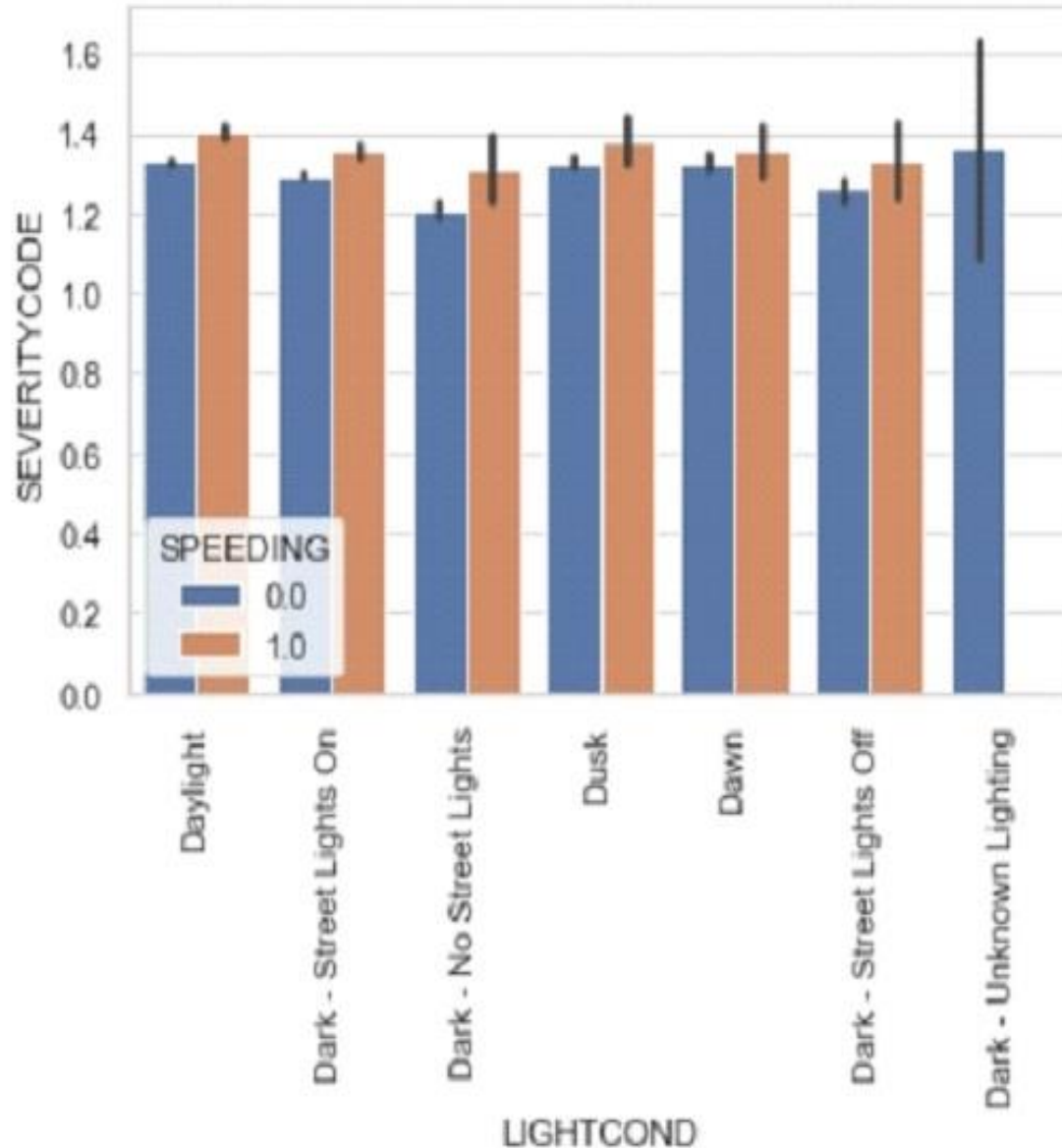
We can look at the
affects of weather
and driver speed
on
"SEVERITYCODE"



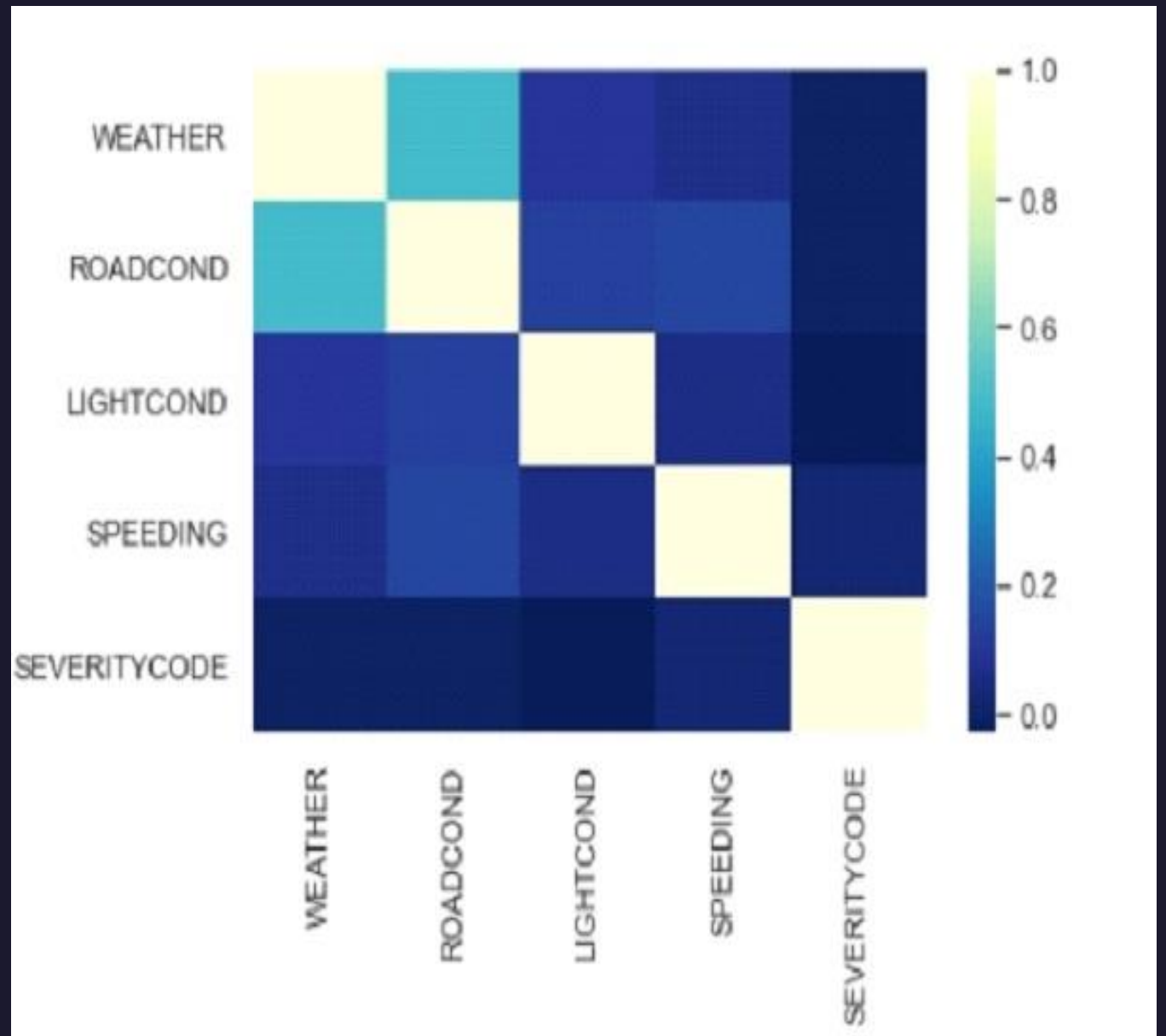
Now we can see
how much ground
condition and
driver speed affect
"SEVERITYCODE"



And finally we can
look at the affects
of ambient light
conditions and
driver speed.



Heatmaps are useful for cross-examining multivariate data by placing variables in rows and columns and coloring cells in the table. Heat maps can be used to show variance between multiple variables, reveal any design, show if any variable is similar to each other, and determine whether there is any correlation between them.



As a result we applied three machine learning algorithms, KNN, Decision Tree, and Logistic regression, on our model. The evaluation criteria used to test the accuracy of our models were the jaccard index and the f-1 score. It can be concluded that certain classes, such as weather conditions, road condition, driver speed, have some effect on injury in vehicle accidents under certain conditions.

	Model	Jaccard	F1
0	KNN	0.662439	0.419013
1	Tree	0.670933	0.401979
2	Log Reg	0.671011	0.401560