Car accident severity

Capstone Project

Week 1

2. Data

2.1 Data Source

The Original Data Came from:

Data-Collisions.csv

3. Feature

**** Create Data Dictionary ****

1. LOCATION : Description of the general location of the collision
2. SEVERITYCODE : A code that corresponds to the severity of the collision:

    * 3 — fatality

    * 2b —serious injury

    * 2 — injury

    * 1 — prop damage

    * 0 — unknown

3. SEVERITYDESC : A detailed description of the severity of the collision
4. COLLISIONTYPE : Collision type
5. PERSONCOUNT : The total number of people involved in the collision
6. PEDCOUNT : The number of pedestrians involved in the collision. This is entered by the state.
7. PEDCYLCOUNT : The number of bicycles involved in the collision. This is entered by the state.
8. VEHCOUNT : The number of vehicles involved in the collision. This is entered by the state.

9. INJURIES : The number of total injuries in the collision. This is entered by the state.
10. SERIOUSINJURIES : The number of serious injuries in the collision. This is entered by the state.
11. FATALITIES : The number of fatalities in the collision. This is entered by the state.
12. INCDATE : The date of the incident.
13. INCDTTM : The date and time of the incident.
14. JUNCTIONTYPE : Category of junction at which collision took place
15. SDOT_COLCODE : A code given to the collision by SDOT.
16. SDOT_COLDESC : A description of the collision corresponding to the collision code.
17. INATTENTIONIND : Whether or not collision was due to inattention. (Y/N)
18. UNDERINFL : Whether or not a driver involved was under the influence of drugs or alcohol.
19. WEATHER : A description of the weather conditions during the time of the collision.
20. ROADCOND : The condition of the road during the collision.
21. LIGHTCOND : The light conditions during the collision.
22. PEDROWNOTGRNT : Whether or not the pedestrian right of way was not granted. (Y/N)
23. SDOTCOLNUM : A number given to the collision by SDOT.
24. SPEEDING : Whether or not speeding was a factor in the collision. (Y/N)
25. ST_COLCODE : A code provided by the state that describes the collision.
26. ST_COLDESC : A description that corresponds to the state's coding designation.
27. SEGLANEKEY : A key for the lane segment in which the collision occurred.
28. CROSSWALKKEY : A key for the crosswalk at which the collision occurred.
29. HITPARKEDCAR : Whether or not the collision involved hitting a parked car. (Y/N)

4.Data Exploration

We look now for the target values (**SEVERITYCODE**) and do some analysis

```
1  df.SEVERITYCODE.value_counts()
```

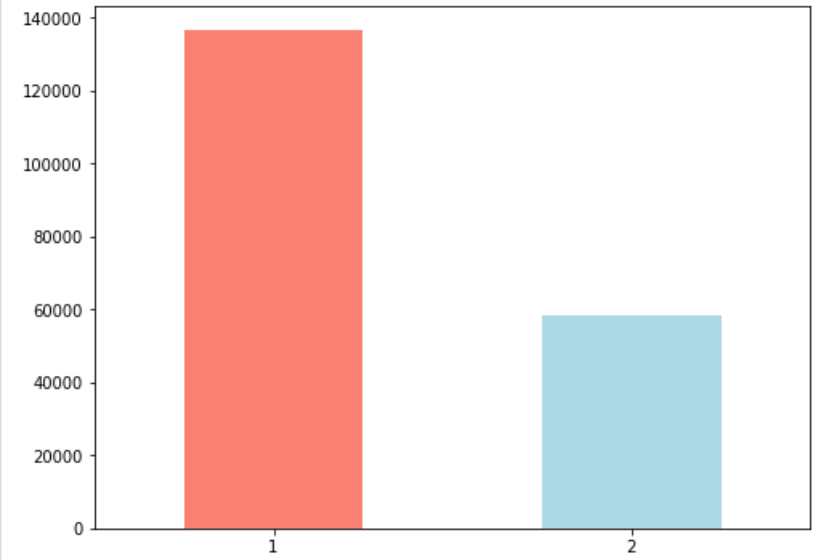
```
1    136485
2     58188
Name: SEVERITYCODE, dtype: int64
```

- We see that most car accident severity :

  1 – prop damage
  2 - injury

- Make some visualization

```
df.SEVERITYCODE.value_counts().plot(kind='bar',figsize=(8,6),color=["salmon", "lightblue"])
plt.xticks(rotation=0);
```

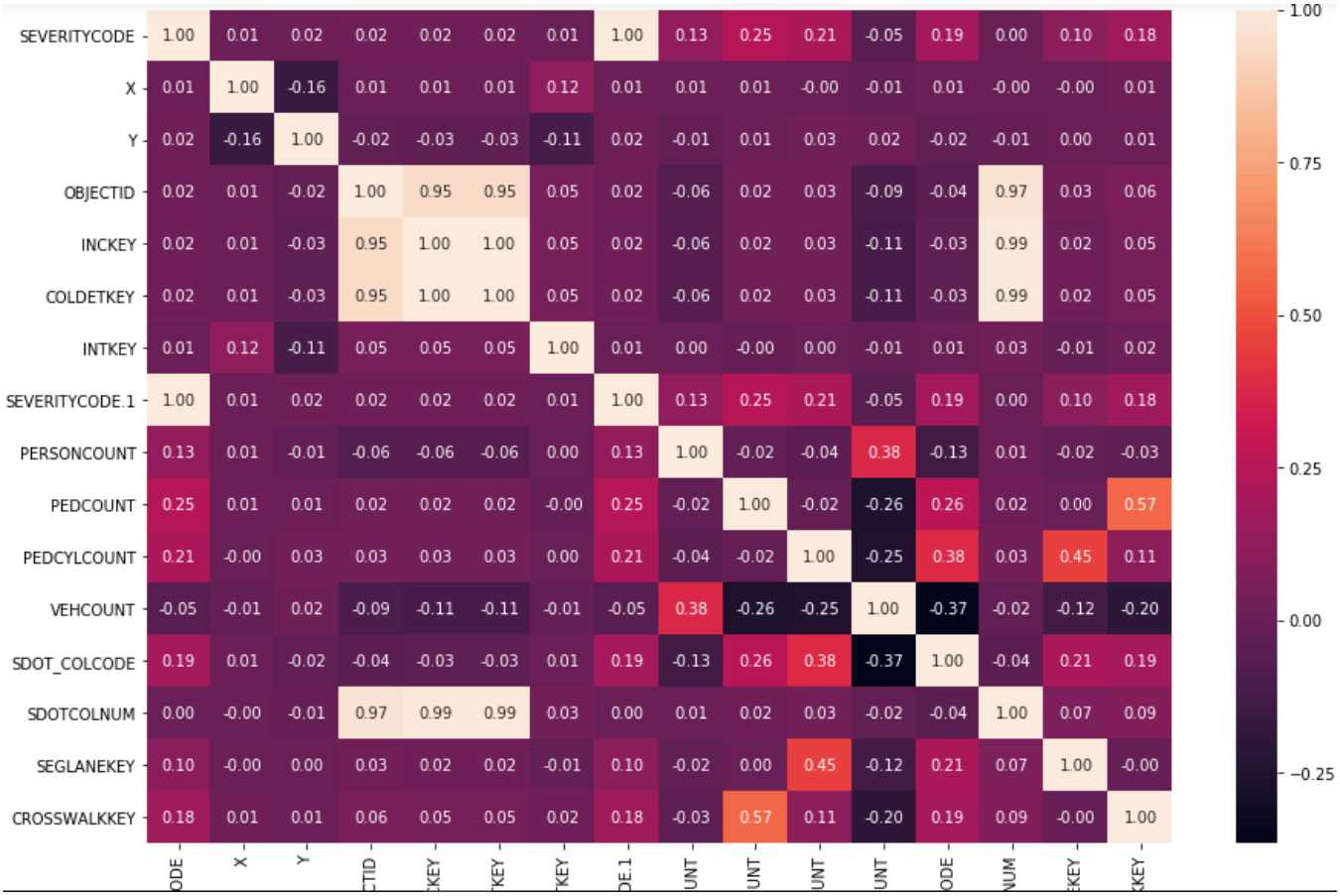


- Here we look to the correlation between Feature

```
# we now see the corrlation beween Features
data_corr = df.corr()
data_corr
```

	SEVERITYCODE	X	Y	OBJECTID	INCKEY	COLDETKEY	INTKEY	SEVERITYCODE.1	PERSONCOUNT	PEDCOUNT	PED(
SEVERITYCODE	1.000000	0.010309	0.017737	0.020131	0.022065	0.022079	0.006553	1.000000	0.130949	0.246338	
X	0.010309	1.000000	-0.160262	0.009956	0.010309	0.010300	0.120754	0.010309	0.012887	0.011304	
Y	0.017737	-0.160262	1.000000	-0.023848	-0.027396	-0.027415	-0.114935	0.017737	-0.013850	0.010178	
OBJECTID	0.020131	0.009956	-0.023848	1.000000	0.946383	0.945837	0.046929	0.020131	-0.062333	0.024604	
INCKEY	0.022065	0.010309	-0.027396	0.946383	1.000000	0.999996	0.048524	0.022065	-0.061500	0.024918	
COLDETKEY	0.022079	0.010300	-0.027415	0.945837	0.999996	1.000000	0.048499	0.022079	-0.061403	0.024914	
INTKEY	0.006553	0.120754	-0.114935	0.046929	0.048524	0.048499	1.000000	0.006553	0.001886	-0.004784	
SEVERITYCODE.1	1.000000	0.010309	0.017737	0.020131	0.022065	0.022079	0.006553	1.000000	0.130949	0.246338	
PERSONCOUNT	0.130949	0.012887	-0.013850	-0.062333	-0.061500	-0.061403	0.001886	0.130949	1.000000	-0.023464	
PEDCOUNT	0.246338	0.011304	0.010178	0.024604	0.024918	0.024914	-0.004784	0.246338	-0.023464	1.000000	
PEDCYLCOUNT	0.214218	-0.001752	0.026304	0.034432	0.031342	0.031296	0.000531	0.214218	-0.038809	-0.016920	
VEHCOUNT	-0.054686	-0.012168	0.017058	-0.094280	-0.107528	-0.107598	-0.012929	-0.054686	0.380523	-0.261285	
SDOT_COLCODE	0.188905	0.010904	-0.019694	-0.037094	-0.027617	-0.027461	0.007114	0.188905	-0.128960	0.260393	
SDOTCOLNUM	0.004226	-0.001016	-0.006958	0.969276	0.990571	0.990571	0.032604	0.004226	0.011784	0.021461	
SEGLANEKEY	0.104276	-0.001618	0.004618	0.028076	0.019701	0.019586	-0.010510	0.104276	-0.021383	0.001810	
CROSSWALKKEY	0.175093	0.013586	0.009508	0.056046	0.048179	0.048063	0.018420	0.175093	-0.032258	0.565326	

- Make Correlation more beautiful



- We now we will select feature that will help us in Machine learning Model

```
#we will focues in some feature that make result
car_acc = df[['WEATHER','ROADCOND','LIGHTCOND','VEHCOUNT','JUNCTIONTYPE','PERSONCOUNT','SEVERITYCODE']]
```

**Data Dictionary **

1. WEATHER : A description of the weather conditions during the time of the collision.
2. ROADCOND : The condition of the road during the collision.
3. LIGHTCOND : The light conditions during the collision.
4. VEHCOUNT : The number of vehicles involved in the collision. This is entered by the state.
5. JUNCTIONTYPE : Category of junction at which collision took place
6. PERSONCOUNT : The total number of people involved in thecollision
7. SEVERITYCODE : A code that corresponds to the severity of the collision:

    * 3 — fatality
    * 2b —serious injury
    * 2 — injury
    * 1 — prop damage
    * 0 — unknown

- This’s our data now

```
1 car_acc
```

	WEATHER	ROADCOND	LIGHTCOND	VEHCOUNT	JUNCTIONTYPE	PERSONCOUNT	SEVERITYCODE
0	Overcast	Wet	Daylight	2	At Intersection (intersection related)	2	2
1	Raining	Wet	Dark - Street Lights On	2	Mid-Block (not related to intersection)	2	1
2	Overcast	Dry	Daylight	3	Mid-Block (not related to intersection)	4	1
3	Clear	Dry	Daylight	3	Mid-Block (not related to intersection)	3	1
4	Raining	Wet	Daylight	2	At Intersection (intersection related)	2	2
...	...	...	...	...	...	...	...
194668	Clear	Dry	Daylight	2	Mid-Block (not related to intersection)	3	2
194669	Raining	Wet	Daylight	2	Mid-Block (not related to intersection)	2	1
194670	Clear	Dry	Daylight	2	At Intersection (intersection related)	3	2
194671	Clear	Dry	Dusk	1	At Intersection (intersection related)	2	2
194672	Clear	Wet	Daylight	2	Mid-Block (not related to intersection)	2	1

194673 rows × 7 columns