Predicting occurrence and severity of traffic accidents

Business Understanding

- All means to prevent accidents should be used
- Several features are available before the trip actually begins in order to assess the risk of getting into a accident
- This work is meant to provide a tool to predict the occurrence and severity of a collision under given circumstances

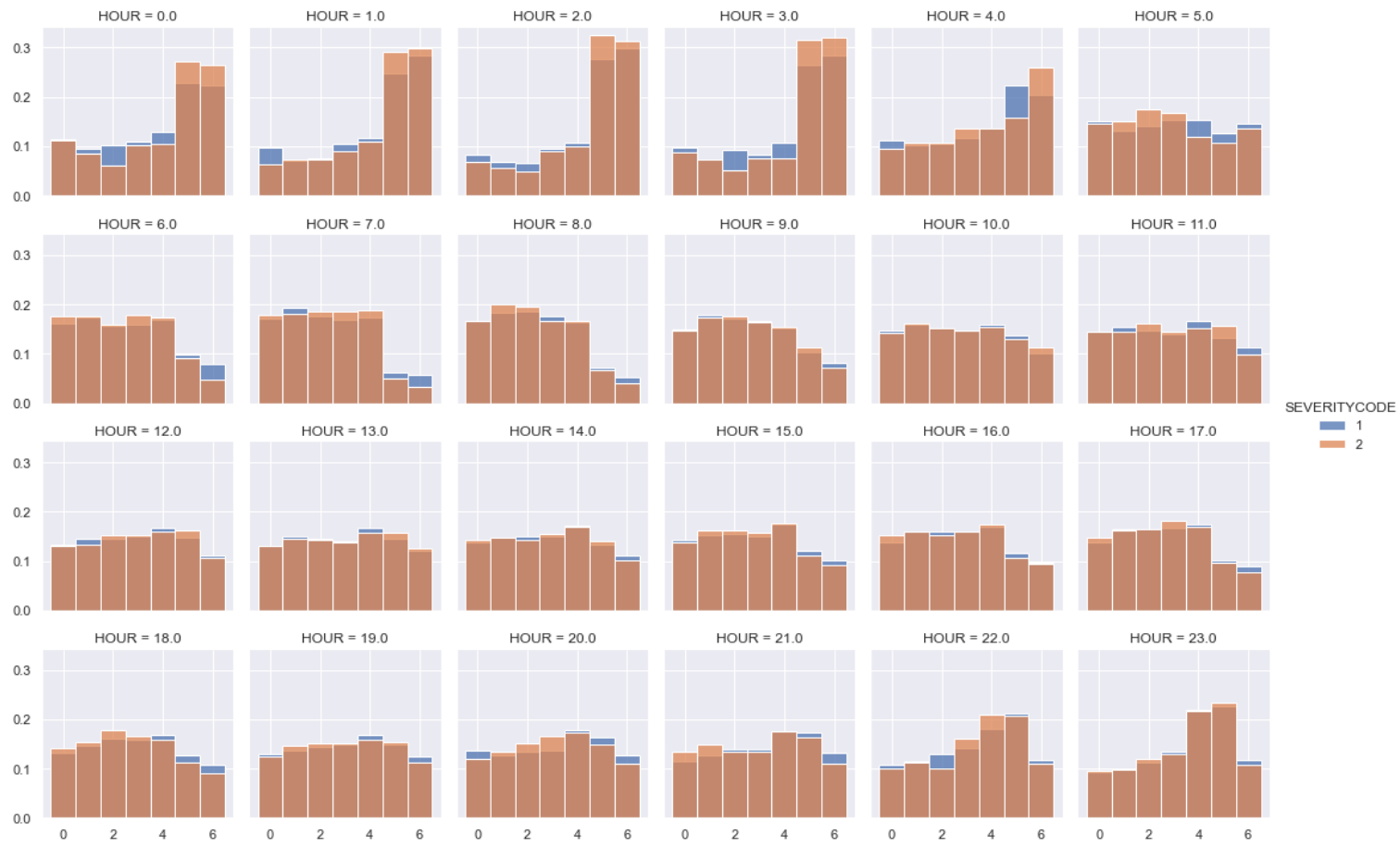
Data Understanding

- dataset provides a collection of all reported collisions in Seattle from 2004 to present
- Original dataset contains 194673 rows and 38 features
- Selection of most important features
  - 'INCDTTM': a time-dependency of probability/severity of accidents is expected
  - 'INATTENTIONIND': the usage of cell phones and also being well rested when driving are important factors governing probability/severity of accidents
  - 'UNDERINFL': consumption of alcohol is one of the main reasons for accidents
  - 'WEATHER': impaired view, e.g. due to havy rain
  - 'ROADCOND': the control over the vehicle can be lost due to bad road conditions
  - 'LIGHTCOND': unfortunate incidence of light could impair sight as well --> probably only matters during daylight
  - 'SPEEDING': by laws of physics, the a collision will be more severe at high velocity

Data Preparation

- Homogenizing labels of each feature
- Deleting rows with missing features
- Encoding categorical features as numerical labels
- Splitting dataset in a training/test dataset in a ration 70/30

- The most significant features to consider seem to be day of week as well as hour:
- lowest occurrence is observed on Sundays between 6 an 9am



Modeling & Evaluation

- KNN classifier with k=6 seems to be most suitable for our purpose

Classifier	f1 score	jaccard score	log loss
KNN (k=8)	0.602575621409465 6	0.659787832536549	N/A
Decision Tree	0.572230314591391 1	0.696708693739920 8	N/A
SVM	0.573403397109297 3	0.696568501607559 8	N/A
Logistic Regression	0.602278927840715 9	0.574206329888006 1	0.696038259720880 2

Conclusion & deployment

- In conclusion, it can be stated that non-urgent trips should be carried out considering following requirements:
  - pay attention to the traffic, i.e. do no use your cell phone, etc.
  - do not be under influence of alcohol, medicaments or drugs while driving
  - adhere to the speed limits
  - drive between 6 to 9am