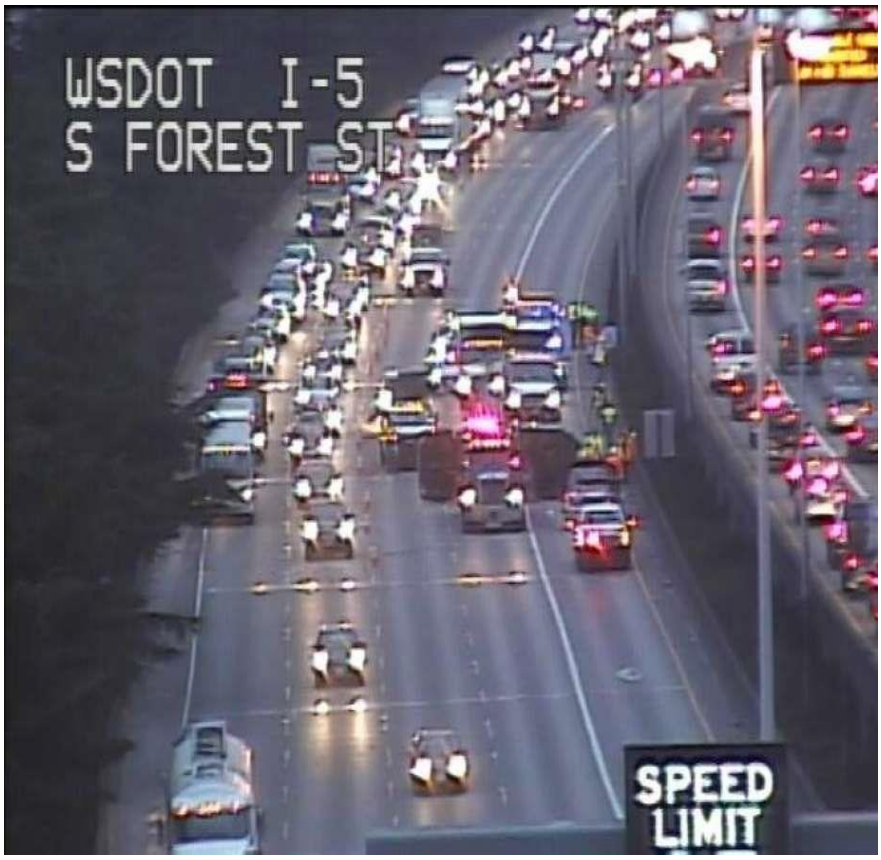

Seattle Occupies List of World's Worst Traffic Cities

(Here a rollover crash blocks all southbound lanes of Interstate 5, causing long traffic backups)

Coursera Data Science Capstone Project - Seattle, Washington, USA Traffic Accident Analysis

Author

Jerry W. Hedgepath

TABLE OF CONTENTS

INTRODUCTION/BUSINESS PROBLEM STATEMENT

SEATTLE TRAFFIC BACKGROUND

According to Joel Connelly, a reporter for the *Seattle Post-Intelligencer*, Seattle ranks as number 20 in a recent CBS News compilation of "Cities with the worst traffic in the world." Seattle is just ahead of Dallas and St Petersburg in Russia, and trails just behind Chicago and Boston. The striking feature of the list is that almost all of the cities on the list are larger than Seattle. The top five cities are: 1) Los Angeles; 2) Moscow; 2 tie) New York City; 4) Sao Paulo, Brazil; and 5) San Francisco, CA.

Seattle's ranking is a product of three "G's – Geography, Growth and Guilt.

Seattle, long known as the Emerald City for lush forests surrounding the city, is sqeezed between two bodies of water, Elliott Bay and Lake Washington. It has just two major north-south highways, Interstate 5 and State Route 99.

It also features world-class examples of engineering ineptitude, such as drivers coming off state Route 520 (the Evergreen Point Bridge), joining southbound I-5 in the left lane, and having less than a mile to cross four lanes of freeway to exit on Mercer. And vice versa.

Seattle has gained more than 100,000 new residents in the past eight years. Cities north, south and east are growing as well.

The guilt? Seattle-area voters twice turned down, in the late 1960's, a proposed rail system. Sen. Warren Magnuson had secured federal money to pay the bulk of the bill. Sadly, the city's construction unions were addicted to concrete, and led the opposition.

On average, Seattle drivers each lost 55 hours to traffic during peak times in 2017.

PROBLEM STATEMENT

With the traffic problems outlined above, the ability to accurately analyze and model traffic accident data becomes increasingly important. A baseline ability to predict the "seriousness" of a future accident is key along with drawing insights into traffic patterns based on time of day, day of week, weather, lighting and road conditions, and other attributes. Additionally, a variety of insights may be derived to benefit urban planning efforts and improving transportation infrastructure.

DATA SOURCES

SEATTLE DEPARTMENT OF TRANSPORTATION TRAFFIC DATA

The homepage of the Seattle Department of Transportation traffic data is:

http://data-seattlecitygis.opendata.arcgis.com/datasets/5b5c745e0f1f48e7a53acec63a0022ab_0.csv

TRAFFIC DATA METADATA

Meta-data of the dataset can be viewed at
https://www.seattle.gov/Documents/Departments/SDOT/GIS/Collisions_OD.pdf

TRAFFIC DATA ANALYSIS

The labelled dataset contains 221,389 data rows. The dataset was last updated on September 5, 2020 and accessed on September 18, 2020. The dataset covers the time frame from 2004 to last update date. The dataset contains 40 attributes some of which may not be useful for modeling.

Some basic information from the dataframe:

a. Head (partial column display)

```
In [4]: #read the data
        pd.set_option('display.max_columns', None)
        df = pd.read_csv("http://data-seattlecitygis.opendata.arcgis.com/datasets/5b5c745e0f1f48e7a53acec63a0022ab_0.csv")
        print("Data read into dataframe!")

        Data read into dataframe!

In [5]: # A cursory analysis of the data
        df.head()

Out[5]:
```

	X	Y	OBJECTID	INCKEY	COLDETKEY	REPORTNO	STATUS	ADDRTYPE	INTKEY	LOCATION	EXCEPTRSNCODE	EXCEPTRSNDESC	SEVERITYCODE	SEVERITYDESC	COLLISIONTYPE	PERSONCOUNT	PEDCOUNT	PEDCYLCOUNT	VEHCOUNT
0	-122.320757	47.609408	1	328476	329976	EA08706	Matched	Block	NaN	BROADWAY BETWEEN E COLUMBIA ST AND BOYLSTON AVE		NaN	1	Property Damage Only Collision	Sideswipe	2	0	0	2
1	-122.319561	47.662221	2	328142	329642	EA06882	Matched	Block	NaN	8TH AVE NE BETWEEN NE 45TH E ST AND NE 47TH ST		NaN	1	Property Damage Only Collision	Parked Car	2	0	0	2
2	-122.327525	47.604393	3	20700	20700	1181833	Unmatched	Block	NaN	JAMES ST BETWEEN 6TH AVE AND 7TH AVE	NaN	NaN	0	Unknown	NaN	0	0	0	0
3	-122.327525	47.708622	4	332126	333626	M16001640	Unmatched	Block	NaN	NE NORTHGATE WAY BETWEEN 1ST AVE NE AND NE NOR...		NaN	0	Unknown	NaN	0	0	0	0
4	-122.292120	47.559009	5	328238	329738	3857118	Unmatched	Block	NaN	M L KING JR ER WAY S BETWEEN S ANGELINE ST AND...		NaN	0	Unknown	NaN	0	0	0	0

b. Shape and column values

```
In [7]: df.shape

Out[7]: (221389, 40)

In [9]: df.columns.values

Out[9]: array(['X', 'Y', 'OBJECTID', 'INCKEY', 'COLDETKEY', 'REPORTNO', 'STATUS',
               'ADDRTYPE', 'INTKEY', 'LOCATION', 'EXCEPTRSNCODE', 'EXCEPTRSNDESC',
               'SEVERITYCODE', 'SEVERITYDESC', 'COLLISIONTYPE', 'PERSONCOUNT',
               'PEDCOUNT', 'PEDCYLCOUNT', 'VEHCOUNT', 'INJURIES',
               'SERIOUSINJURIES', 'FATALITIES', 'INCDATE', 'INCDTTM',
               'JUNCTIONTYPE', 'SDOT_COLCODE', 'SDOT_COLDESC', 'INATTENTIONIND',
               'UNDERINFL', 'WEATHER', 'ROADCOND', 'LIGHTCOND', 'PEDROWNOTGRNT',
               'SDOTCOLNUM', 'SPEEDING', 'ST_COLCODE', 'ST_COLDESC', 'SEGLANEKEY',
               'CROSSWALKKEY', 'HITPARKEDCAR'], dtype=object)
```

c. Datatypes

```
In [12]: df.dtypes
```

```
Out[12]: X                 float64
         Y                 float64
         OBJECTID            int64
         INCKEY              int64
         COLDETKEY           int64
         REPORTNO           object
         STATUS             object
         ADDRTYPE           object
         INTKEY            float64
         LOCATION           object
         EXCEPTRSNCODE      object
         EXCEPTRSNDESC      object
         SEVERITYCODE       object
         SEVERITYDESC       object
         COLLISIONTYPE      object
         PERSONCOUNT         int64
         PEDCOUNT            int64
         PEDCYLCOUNT         int64
         VEHCOUNT            int64
         INJURIES            int64
         SERIOUSINJURIES     int64
         FATALITIES          int64
         INCDATE            object
         INCDTTM            object
         JUNCTIONTYPE       object
         SDOT_COLCODE      float64
         SDOT_COLDESC       object
         INATTENTIONIND     object
         UNDERINFL          object
         WEATHER            object
         ROADCOND           object
         LIGHTCOND          object
         PEDROWNOTGRNT      object
         SDOTCOLNUM        float64
         SPEEDING           object
         ST_COLCODE         object
         ST_COLDESC         object
         SEGLANEKEY          int64
         CROSSWALKKEY        int64
         HITPARKEDCAR       object
         dtype: object
```

d. Datatypes counts

```
In [11]: df.dtypes.value_counts()

Out[11]: object     23
         int64      12
         float64     5
         dtype: int64
```

METHODOLOGY SECTION

RESULTS SECTION

DISCUSSION SECTION

CONCLUSION SECTION