

# Capstone Project

Applied Data Science Capstone by IBM/Coursera

## Table of contents

- [Introduction: Business Problem](#)
- [Data](#)
- [Methodology](#)
- [Analysis](#)
- [Results and Discussion](#)
- [Conclusion](#)



# Introduction Business Problem

## English

With traffic accidents that claim more than 40,000 lives per year in the United States, **It is necessary to investigate the causes that generate, focusing on finding a solution that we can take to avoid accidents or reduce their severity by analyzing data from the Seattle Department of Transportation.**

Seattle is the largest city in the state of Washington, in the northwestern United States of America. The metropolitan area of the city includes Seattle-Tacoma-Bellevue, is the 15th most populous in the country and the largest in the Pacific Northwest. The city is located between Lake Washington and the bay known as Puget Sound, along the Pacific Ocean. It is located 155 kilometers south of the border between the United States and Canada. A major cultural, educational, and economic center of the region, Seattle is the seat of King County. In the estimated 2018 census, the city had a municipal population of 744,955 inhabitants, making it the 24th most populous city in the United States, and with a metropolitan area of 3,263,497 inhabitants.

After the economic recession between 2007 and 2009, Seattle gained an average of almost 15,000 residents per year for the next 5 years. When Amazon.com moved its headquarters to South Lake Union, a construction boom began, resulting in the completion of nearly 10,000 Seattle apartments in 2017 (more than any previous year and twice as many as in 2016). Although unemployment fell from 9% to 3.6%, the city found itself overflowing and had the sixth worst peak hour traffic.

Seattle's climate is mild, with a moderate temperature by the sea, protected from winds and storms by the mountains. The city of Seattle has a reputation for frequent rains, although "the rainy city" only receives about 970 mm of precipitation per year, less than almost all the major cities on the east coast of the United States, such as the city of New York, which receives on average 1200 mm. Although it rains much more than in other places in the world such as Madrid (430 mm) or Mexico City (830 mm). The worldwide fame of the rains in Seattle is due to the fact that, like in London, almost all of its precipitation falls in the form of drizzle or light rain, since Seattle is in the orographic pluviometric shadow of the Olympic Mountains. In other words, although it rains regularly, it does not rain very hard. Most of the rains fall from November to March. December is the雨iest month.

In 2000, the state of Washington adopted "Target Zero," a plan aimed at ending traffic fatalities and serious injuries by 2030. The drive for safer streets expanded to Seattle's "Vision Zero" in 2015 , after collaborating with the state and achieving a 28% reduction in fatalities and serious injuries at Aurora Avenue North. Originally implemented in Sweden in the 1990s, Vision Zero has proven to be a success across Europe, and many cities in the US have signed on to the idea that even a single death from these accidents is unacceptable and preventable. This study is expected to reveal what steps, if any, we can take as individuals and municipalities to make traveling in Seattle safer.

## Español

Con los accidentes de tráfico que cobran más de 40.000 vidas por año en los EE. UU , **Es necesario investigar las causa que generar, poniendo el foco en buscar una solucion que podamos tomar para evitar accidentes o reducir su gravedad mediante el análisis de datos del Departamento de Transporte de Seattle.**

Seattle la ciudad más grande del estado de Washington, en el noroeste de los Estados Unidos de América. El área metropolitana de la ciudad comprende Seattle-Tacoma-Bellevue, es la 15.<sup>a</sup> más poblada del país y la mayor del Noroeste del Pacífico. La ciudad está situada entre el lago Washington y la bahía conocida como Puget Sound, junto al océano Pacífico. Se encuentra a 155 kilómetros al sur de la frontera entre Estados Unidos y Canadá. Importante centro cultural, educativo y económico de la región, Seattle es sede del

condado de King. En el censo estimado de 2018, la ciudad contaba con una población municipal de 744.955 habitantes, por lo que es la vigésimo cuarta ciudad más poblada de los Estados Unidos, y con un área metropolitana de 3.263.497 habitantes.

Después de la recesión económica entre 2007 y 2009, Seattle ganó un promedio de casi 15.000 residentes por año durante los próximos 5 años. Cuando Amazon.com trasladó su sede a South Lake Union, comenzó un auge de la construcción, que resultó en la finalización de casi 10,000 apartamentos en Seattle en 2017 (más que cualquier año anterior y el doble que en 2016). Aunque el desempleo bajó del 9% al 3,6%, la ciudad se encontró a rebosar y tuvo el sexto peor tráfico en las horas pico.

El clima de Seattle es suave, con una temperatura moderada por el mar, protegida de vientos y tormentas por las montañas. La ciudad de Seattle tiene una reputación por sus lluvias frecuentes, aunque "la ciudad lluviosa" solo recibe unos 970 mm de precipitación por año, menos que casi todas las ciudades mayores de la costa este de los Estados Unidos, como por ejemplo la ciudad de Nueva York, que recibe en promedio 1200 mm. Aunque llueve mucho más que en otros lugares del mundo como Madrid (430 mm) o Ciudad de México (830 mm). La fama mundial de las lluvias en Seattle es debido al hecho de que, al igual que en Londres, casi toda su precipitación cae en la forma de llovizna o lluvia ligera, dado que Seattle está en la sombra orográfica pluviométrica de los montes Olympic. O sea, aunque llueve con regularidad, no llueve muy fuerte. La mayoría de las lluvias caen de noviembre a marzo. Diciembre es el mes más lluvioso.

En 2000, el estado de Washington adoptó "Target Zero", un plan destinado a poner fin a las muertes por accidentes de tránsito y las lesiones graves para 2030. El impulso por calles más seguras se expandió a "Vision Zero" de Seattle en 2015, después de colaborar con el estado y lograr una reducción del 28% en muertes y lesiones graves en Aurora Avenue North. Originalmente implementado en Suecia en la década de 1990, Vision Zero ha demostrado ser un éxito en toda Europa, y muchas ciudades de los EE. UU. Han firmado la idea de que incluso una sola muerte causada por estos accidentes es inaceptable y prevenible. Se espera que este estudio revele qué medidas, si las hay, podemos tomar como individuos y municipios para hacer que viajar en Seattle sea más seguro.

## 2. Data

### English

#### **Los datos utilizados para este estudio se obtuvieron del Departamento de Transporte de Seattle.**

The data used is from Seattle's Collision GIS (Geographic Information System), a computer system used for capturing, storing and displaying data related to positions on Earth's surface. The data is from 2004 to the present and contains various features such as location, the severity of the collision, number of vehicles/cyclists/pedestrians involved, date/time of incident, weather, road conditions and more. There are almost 200,000 collisions in the dataset and 38 features. While some of the features won't be useful or have many missing values, the ones that I will explore in more detail will be:

- Severity Code - this will be the target that we'll compare the features' impact on.
- Severity Description - description of the severity codes.
- X and Y values (coordinates) - are there areas where collisions are more concentrated?
- Address Type - alley, block or intersection of collision.
- Collision Type - 10 types of collisions such as parked car, angles, rear end, pedestrian, etc.
- Person Count - # of people involved in collision.
- Pedestrian Count - # of pedestrians involved in collision.
- Cyclist Count - # of cyclists involved in collision.
- Vehicle count - # of vehicles involved in collision.
- Date/Time - are number of or severity of collisions more likely to occur on certain days or times?
- Junction Type - 7 types describing collision at intersection, mid-block, driveway and whether collision is related to intersection.
- Seattle Collision Code - Seattle codes to describe each collision.
- Seattle Collision Description - description of Seattle collision codes.
- Under Influence - was alcohol or drugs involved?
- Weather - do more collisions occur because of adverse weather?
- Road Conditions - do more collisions occur because of adverse road conditions?
- Light Conditions - do more collisions occur because of adverse light conditions?
- State Collision Code - 84 codes the state uses to describe each collision.
- State Collision Description - description of state collision codes.
- Hit Parked Car - was a parked car involved in the collision?

In the following section I will use graphs to gain insight into the data, and decide which features will be useful when predicting injuries when these collisions occur.

### Español

#### **Los datos utilizados para este estudio se obtuvieron del Departamento de Transporte de Seattle.**

Los datos utilizados proceden del Collision GIS (Sistema de Información Geográfica) de Seattle, un sistema informático utilizado para capturar, almacenar y mostrar datos relacionados con las posiciones en la superficie de la Tierra. Los datos son desde 2004 hasta el presente y contienen varias características como la ubicación, la gravedad de la colisión, el número de vehículos / ciclistas / peatones involucrados, la fecha / hora del incidente, el clima, las condiciones de la carretera y más. Hay casi 200.000 colisiones en el conjunto de datos y 38 características. Si bien algunas de las características no serán útiles o tendrán muchos valores faltantes, las que exploraré con más detalle serán:

- Código de gravedad: este será el objetivo en el que compararemos el impacto de las funciones.

- Descripción de gravedad: descripción de los códigos de gravedad.
- Valores X e Y (coordenadas): ¿hay áreas donde las colisiones están más concentradas?
- Tipo de dirección: callejón, cuadra o intersección de colisión.
- Tipo de colisión: 10 tipos de colisiones, como automóvil estacionado, ángulos, parte trasera, peatón, etc.
- Recuento de personas: número de personas involucradas en la colisión.
- Recuento de peatones: número de peatones involucrados en una colisión.
- Conteo de ciclistas: número de ciclistas involucrados en una colisión.
- Recuento de vehículos: número de vehículos involucrados en la colisión.
- Fecha / hora: ¿es más probable que ocurra el número o la gravedad de las colisiones en determinados días u horas?
- Tipo de cruce: 7 tipos que describen la colisión en la intersección, la mitad de la cuadra, el camino de entrada y si la colisión está relacionada con la intersección.
- Código de colisión de Seattle: códigos de Seattle para describir cada colisión.
- Descripción de colisión de Seattle: descripción de los códigos de colisión de Seattle.
- Bajo influencia: ¿hubo alcohol o drogas involucrados?
- Clima: ¿ocurren más colisiones debido al clima adverso?
- Condiciones de la carretera: ¿se producen más colisiones debido a condiciones adversas de la carretera?
- Condiciones de luz: ¿se producen más colisiones debido a condiciones de luz adversas?
- Código de colisión estatal: 84 códigos que el estado usa para describir cada colisión.
- Descripción de colisión de estado: descripción de los códigos de colisión de estado.
- Coche estacionado atropellado: ¿hubo un automóvil estacionado involucrado en la colisión?

En la siguiente sección, usaré gráficos para comprender mejor los datos y decidir qué características serán útiles para predecir lesiones cuando ocurran estas colisiones.

### 3. Metodology

#### English

Jupyter Notebooks was used to perform the analysis and all the necessary Python libraries such as Pandas, Numpy, Matplotlib and Seaborn will be imported. The data was mostly categorical, therefore most variables are plotted in graphs to see the correlation between various variables.

The csv file was imported and I prepare the data, I dropped the columns that we do not need from the dataset, that is, columns that have no values or where the values are unknown.

#### Español

Se utilizo Jupyter Notebooks para realizar el análisisy se importo todas las bibliotecas de Python necesarias como Pandas, Numpy, Matplotlib y Seaborn. Los datos eran en su mayoría categóricos, por ende se represento la mayoria de las variables en graficos para ver la correlación entre varias variables.

Se importo el archivo csv y preparo los datos, dejé caer las columnas que no necesitamos del conjunto de datos, es decir, columnas que no tienen valores o donde los valores son desconocidos.

### 4. Analysis

In [3]:

```
# Import Libraries

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import matplotlib as mpl
import seaborn as sns
%matplotlib inline
```

In [52]:

```
# Load data into datafram and Look at first 5 collisions
pd.set_option('display.max_columns', None)
df = pd.read_csv('https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Data-Collisions.csv', low_memory=False)
df.head()
```

Out[52]:

	SEVERITYCODE	X	Y	OBJECTID	INCKEY	COLDETKEY	REPORTNO
0	2	-122.323148	47.703140		1	1307	1307
1	1	-122.347294	47.647172		2	52200	52200
2	1	-122.334540	47.607871		3	26700	26700
3	1	-122.334803	47.604803		4	1144	1144
4	2	-122.306426	47.545739		5	17700	17700

In [53]:

```
print('Number of Features: ', df.shape[1])
print('Number of Recorded Collisions: ', df.shape[0])
```

Number of Features: 38  
 Number of Recorded Collisions: 194673

In [54]:

```
# check number of missing values for each feature  
df.isnull().sum()
```

Out[54]:

```
SEVERITYCODE          0  
X                  5334  
Y                  5334  
OBJECTID            0  
INCKEY              0  
COLDETKEY          0  
REPORTNO            0  
STATUS              0  
ADDRTYPE           1926  
INTKEY             129603  
LOCATION            2677  
EXCEPTRSNCODE      109862  
EXCEPTRSNDESC       189035  
SEVERITYCODE.1        0  
SEVERITYDESC         0  
COLLISIONTYPE       4904  
PERSONCOUNT          0  
PEDCOUNT             0  
PEDCYLCOUNT          0  
VEHCOUNT             0  
INCDATE              0  
INCDTTM              0  
JUNCTIONTYPE        6329  
SDOT_COLCODE          0  
SDOT_COLDESC          0  
INATTENTIONIND      164868  
UNDERINFL            4884  
WEATHER              5081  
ROADCOND              5012  
LIGHTCOND             5170  
PEDROWNOTGRNT       190006  
SDOTCOLNUM           79737  
SPEEDING             185340  
ST_COLCODE            18  
ST_COLDESC            4904  
SEGLANEKEY            0  
CROSSWALKKEY          0  
HITPARKEDCAR          0  
dtype: int64
```

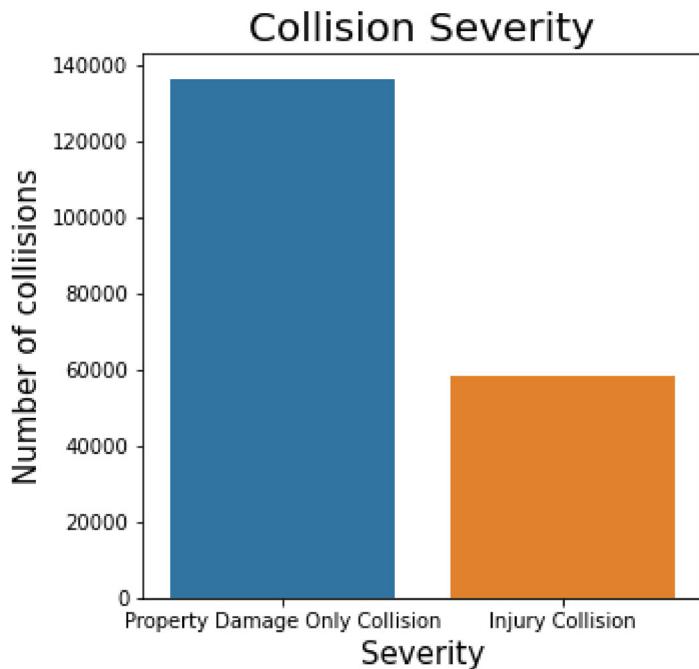
In [55]:

```
def plot_bar(data_column, x_label, y_label, title, size=(12,8), rotate=0):
    a = data_column.value_counts()
    codes = a.index
    count = a.values
    plt.figure(figsize=size)
    ax = sns.barplot( x = codes, y = count )
    ax.set_xlabel(x_label, fontsize=15)
    ax.set_ylabel(y_label, fontsize=15)
    ax.set_title(title, fontsize=20)
    plt.xticks(rotation=rotate)
    plt.show()

def plot_barh(data_column, x_label, y_label, title, size=(12,8), rotate=0):
    a = data_column.value_counts()
    codes = a.index
    count = a.values
    plt.figure(figsize=size)
    ax = sns.barplot( x = count, y = codes )
    ax.set_xlabel(x_label, fontsize=15)
    ax.set_ylabel(y_label, fontsize=15)
    ax.set_title(title, fontsize=20)
    plt.xticks(rotation=rotate)
    plt.show()
```

In [8]:

```
plot_bar(df[ 'SEVERITYDESC' ], 'Severity', 'Number of collisions', 'Collision Severity', size=(5,5))
```



In [9]:

```
!conda install -c conda-forge folium=0.5.0 --yes
import folium
```

```
Collecting package metadata (current_repodata.json): ...working... done
Solving environment: ...working... failed with initial frozen solve. Retry
ing with flexible solve.
Collecting package metadata (repodata.json): ...working... done
Solving environment: ...working... done
```

```
## Package Plan ##
```

```
environment location: C:\Users\nicof\anaconda3
```

```
added / updated specs:
- folium=0.5.0
```

The following packages will be DOWNGRADED:

folium	0.11.0-py_0 --> 0.5.0-py_0
--------	----------------------------

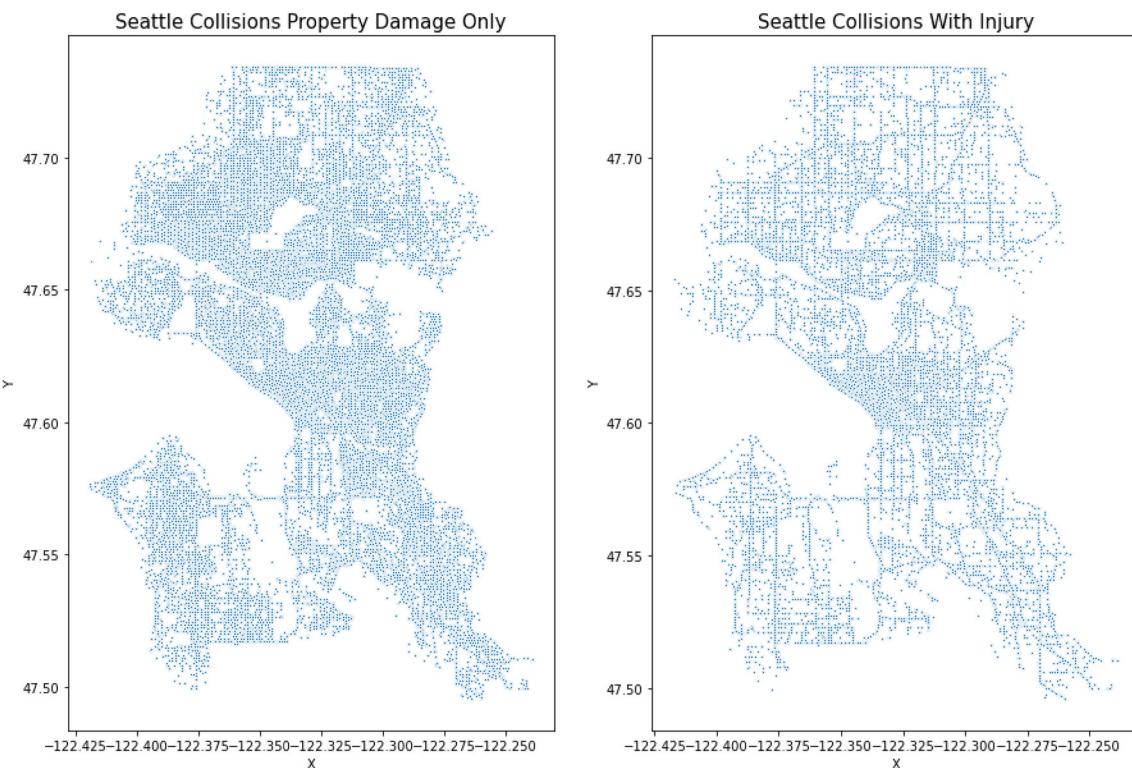
```
Preparing transaction: ...working... done
Verifying transaction: ...working... done
Executing transaction: ...working... done
```

In [10]:

```
# Grafico en el mapa
```

```
property_damage_df = df.loc[df['SEVERITYCODE']==1]
injury_df = df.loc[df['SEVERITYCODE']==2]
fig, (ax1, ax2) = plt.subplots(1,2, figsize=(15,10))
fig.suptitle('Collisions With and Without Injury', fontsize=20)
ax1 = sns.scatterplot(x=property_damage_df['X'], y=property_damage_df['Y'], s=5, ax=ax1)
ax1.set_title('Seattle Collisions Property Damage Only', fontsize=15)
ax2 = sns.scatterplot(x=injury_df['X'], y=injury_df['Y'], s=5, ax=ax2)
ax2.set_title('Seattle Collisions With Injury', fontsize=15);
```

Collisions With and Without Injury



In [11]:

```
#coordenadas seattle
longitud_seattle = -122.3320700
latitud_seattle = 47.6062100
seattle_map = folium.Map(location=[latitud_seattle, longitud_seattle], zoom_start=11)
```

In [12]:

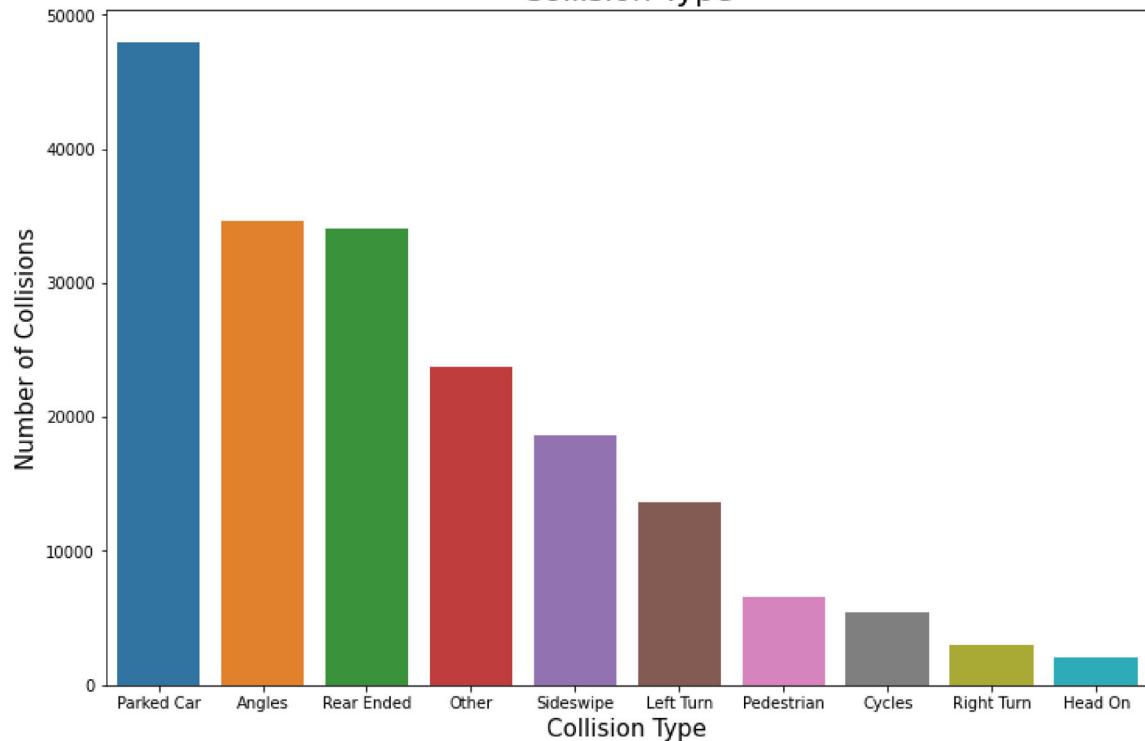
```
print('Cantidad de NAN: ',property_damage_df['X'].isnull().sum())
print('Cantidad de NAN: ',property_damage_df['Y'].isnull().sum())
print('Cantidad de NAN: ',property_damage_df['SEVERITYCODE'].isnull().sum())
property_damage_df = property_damage_df.dropna(axis=0, subset=['X','Y','SEVERITYCODE'])
```

Cantidad de NAN: 4264  
Cantidad de NAN: 4264  
Cantidad de NAN: 0

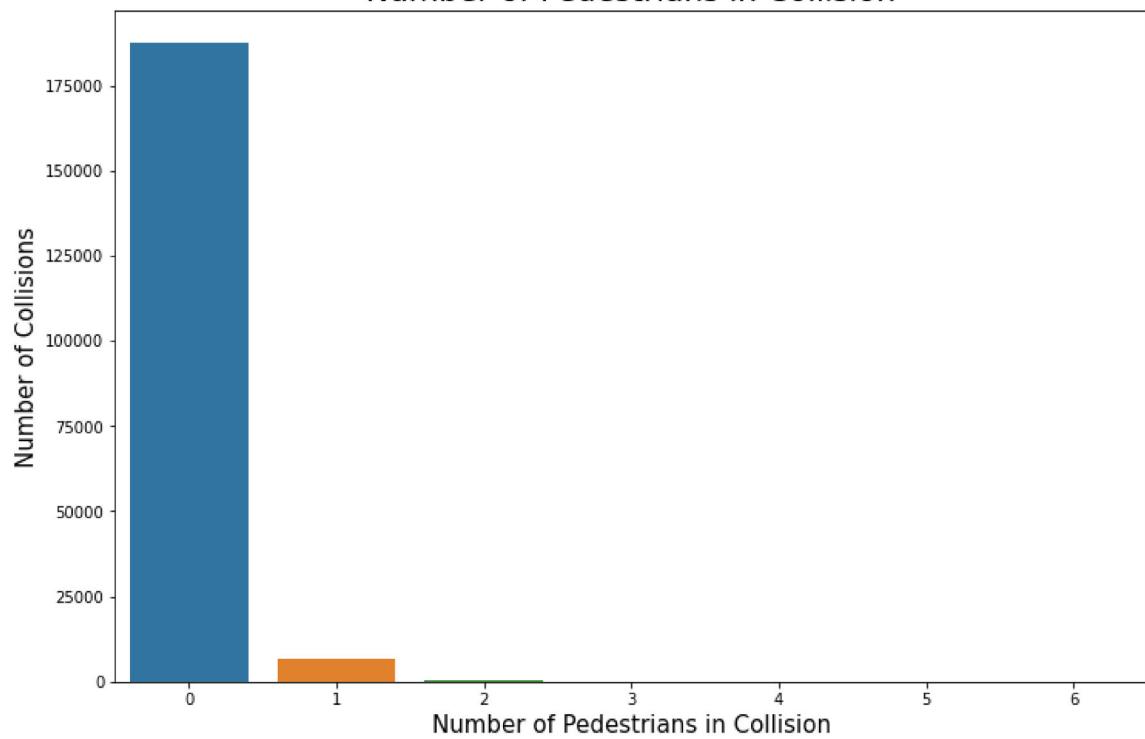
In [13]:

```
plot_bar(df['COLLISIONTYPE'], 'Collision Type', 'Number of Collisions', 'Collision Type', rotate=0)
plot_bar(df['PEDCOUNT'], 'Number of Pedestrians in Collision', 'Number of Collisions', 'Number of Pedestrians in Collision')
```

### Collision Type

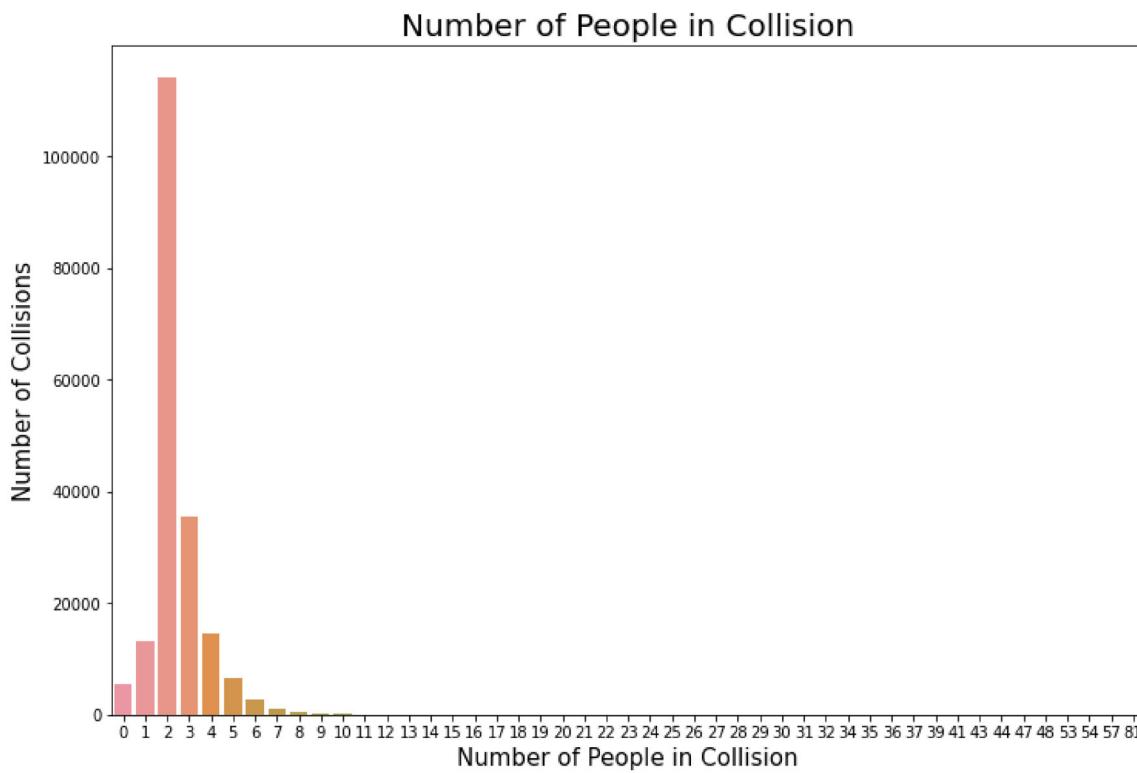
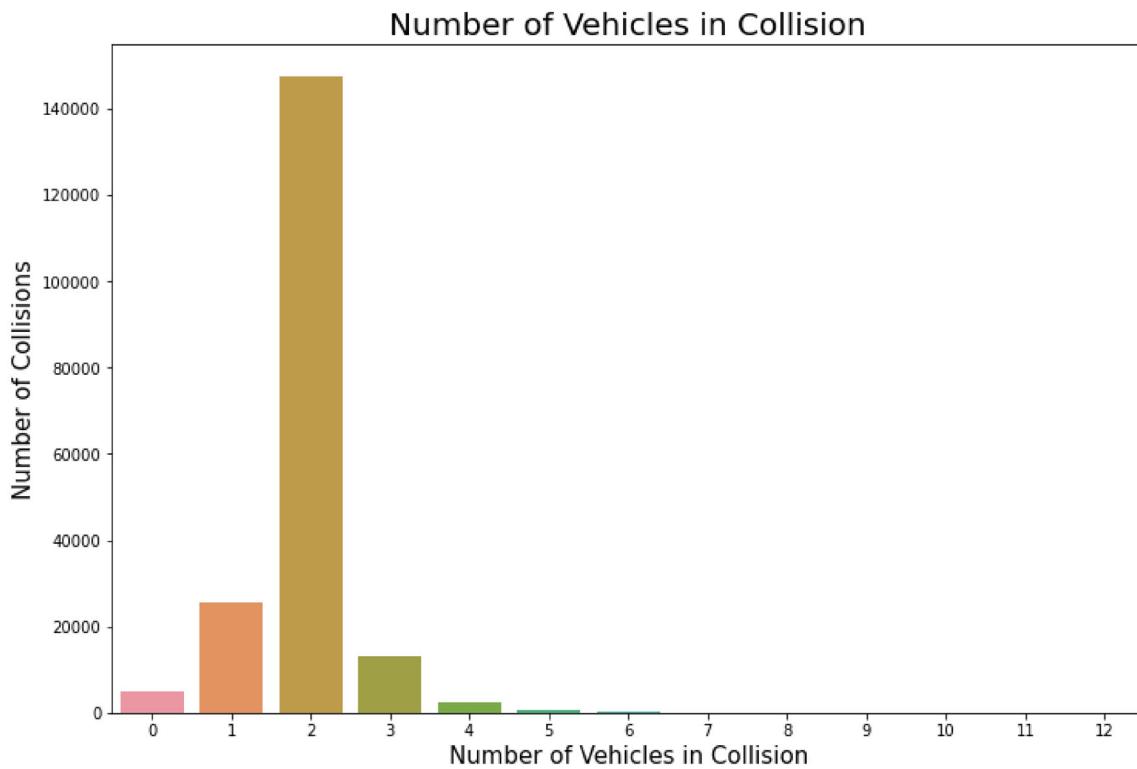


### Number of Pedestrians in Collision



In [14]:

```
plot_bar(df['VEHCOUNT'], 'Number of Vehicles in Collision', 'Number of Collisions', 'Nu  
mber of Vehicles in Collision')  
plot_bar(df['PERSONCOUNT'], 'Number of People in Collision', 'Number of Collisions', 'N  
umber of People in Collision')
```



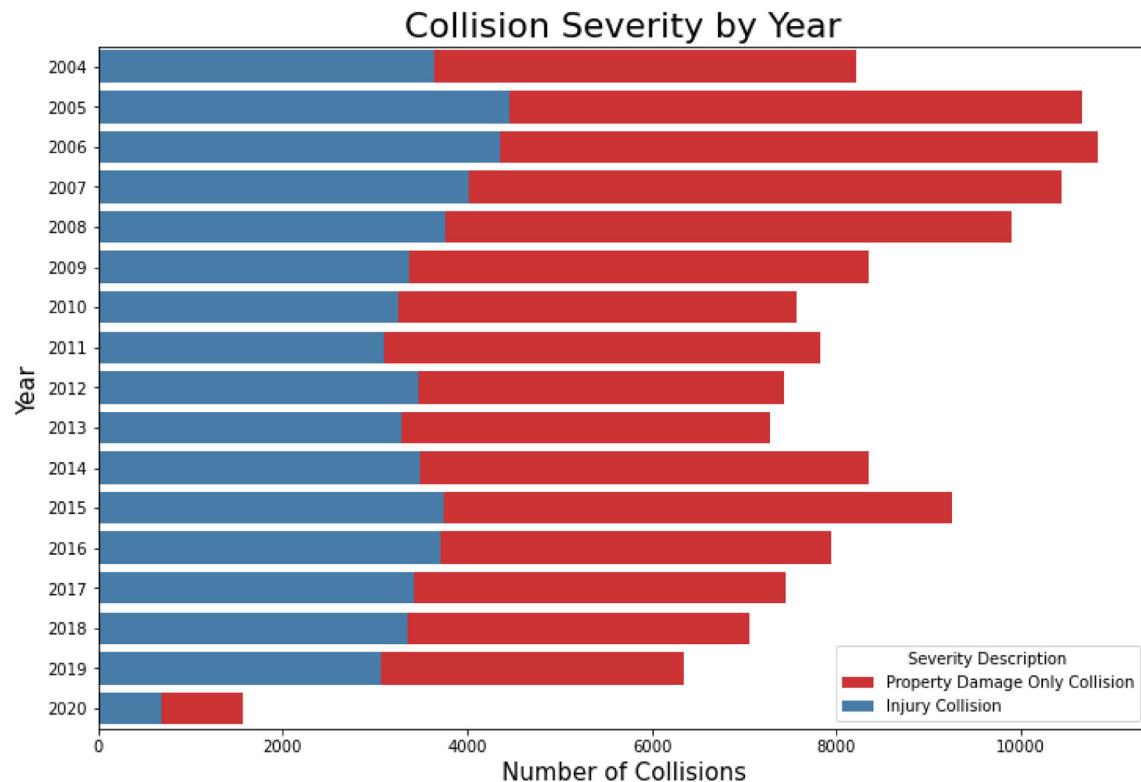
In [15]:

```
#Normalizando fecha de accidente
```

```
df['INCDTTM'] = pd.to_datetime(df['INCDTTM'], infer_datetime_format=True)
df['Year'] = df['INCDTTM'].dt.year
df['Month'] = df['INCDTTM'].dt.month
df['Day'] = df['INCDTTM'].dt.weekday
```

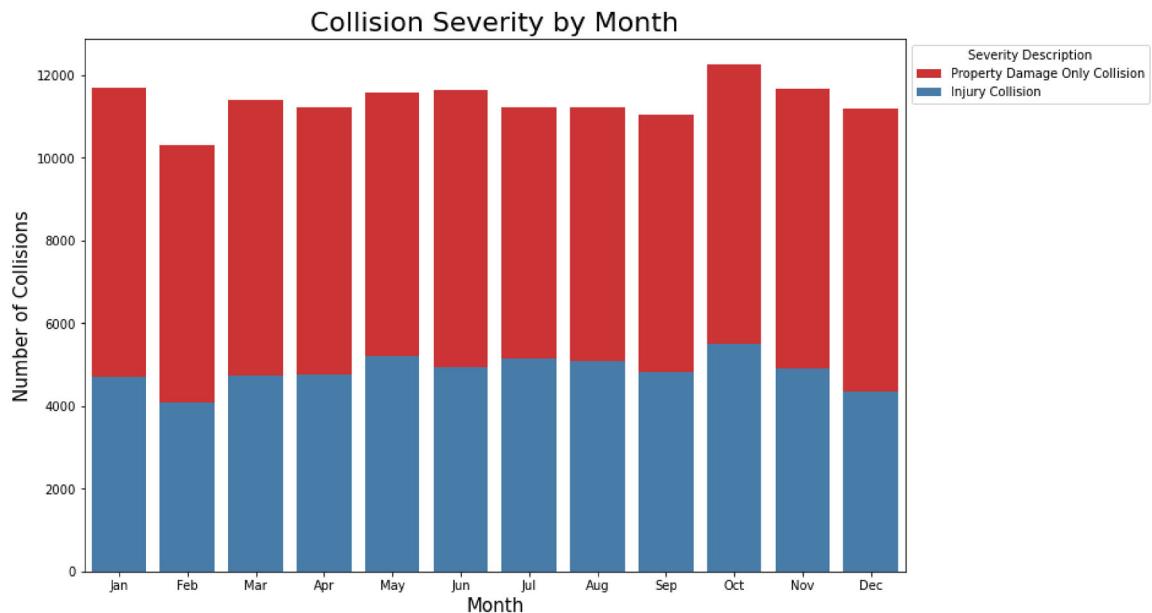
In [16]:

```
plt.figure(figsize=(12,8))
ax = sns.countplot(y='Year', hue='SEVERITYDESC', data=df,
                    hue_order=['Property Damage Only Collision', 'Injury Collision'], dodge=False, palette="Set1")
ax.set_ylabel('Year', fontsize=15)
ax.set_xlabel('Number of Collisions', fontsize=15)
ax.set_title('Collision Severity by Year', fontsize=22)
ax.legend(title='Severity Description');
```



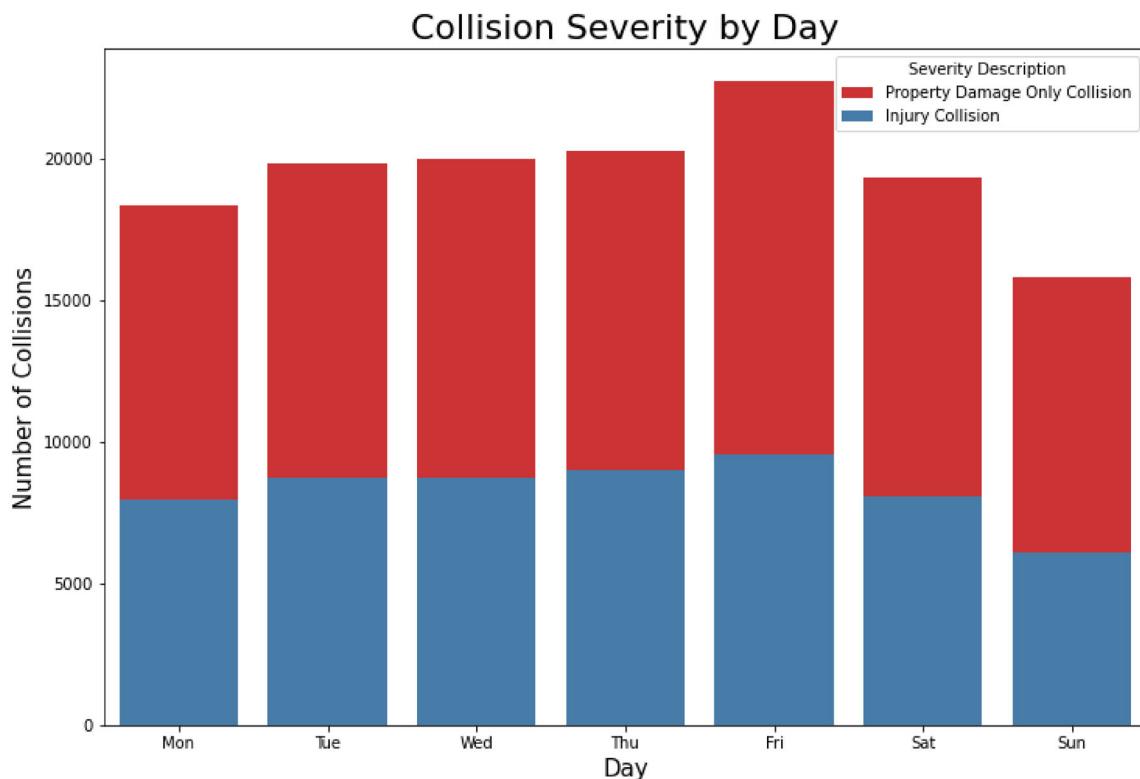
In [17]:

```
months=['Jan','Feb','Mar','Apr','May','Jun','Jul','Aug','Sep','Oct','Nov','Dec']
plt.figure(figsize=(12,8))
ax = sns.countplot(x='Month', hue='SEVERITYDESC', data=df,
                    hue_order=[ 'Property Damage Only Collision', 'Injury Collision'], do
dge=False, palette="Set1")
ax.set_xlabel('Month', fontsize=15)
ax.set_ylabel('Number of Collisions', fontsize=15)
ax.set_title('Collision Severity by Month', fontsize=22)
ax.set_xticklabels(months)
ax.legend(title='Severity Description', bbox_to_anchor=(1,1));
```



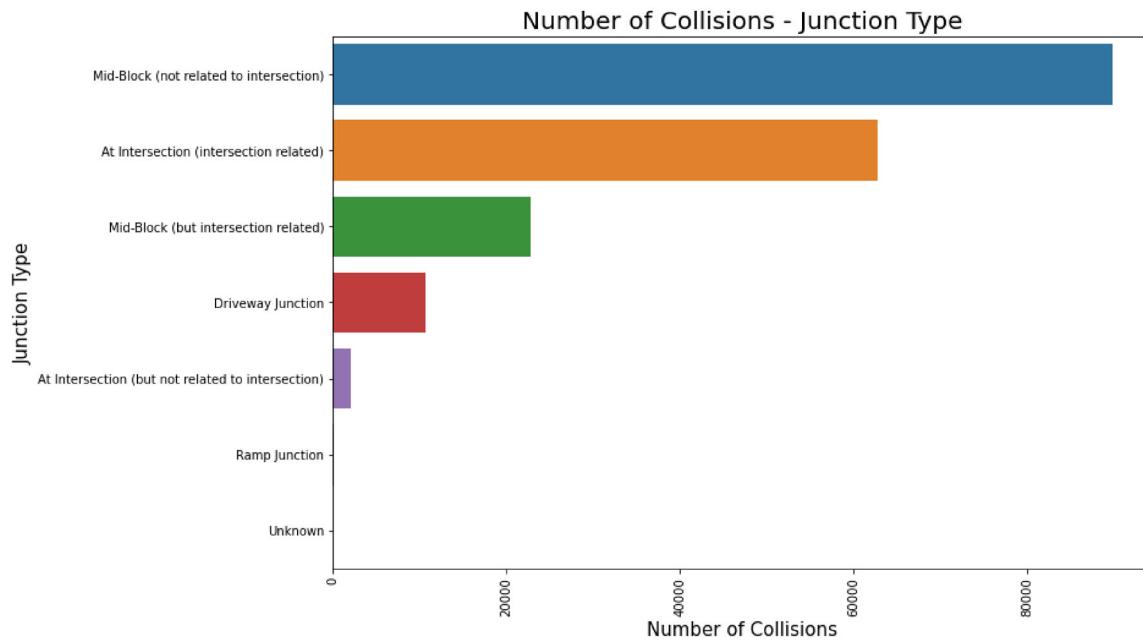
In [18]:

```
days=['Mon','Tue','Wed','Thu','Fri','Sat','Sun']
plt.figure(figsize=(12,8))
ax = sns.countplot(x='Day', hue='SEVERITYDESC', data=df,
                    hue_order=[ 'Property Damage Only Collision', 'Injury Collision'], do
dge=False, palette="Set1")
ax.set_xlabel('Day', fontsize=15)
ax.set_ylabel('Number of Collisions', fontsize=15)
ax.set_title('Collision Severity by Day', fontsize=22)
ax.set_xticklabels(days)
ax.legend(title='Severity Description');
```



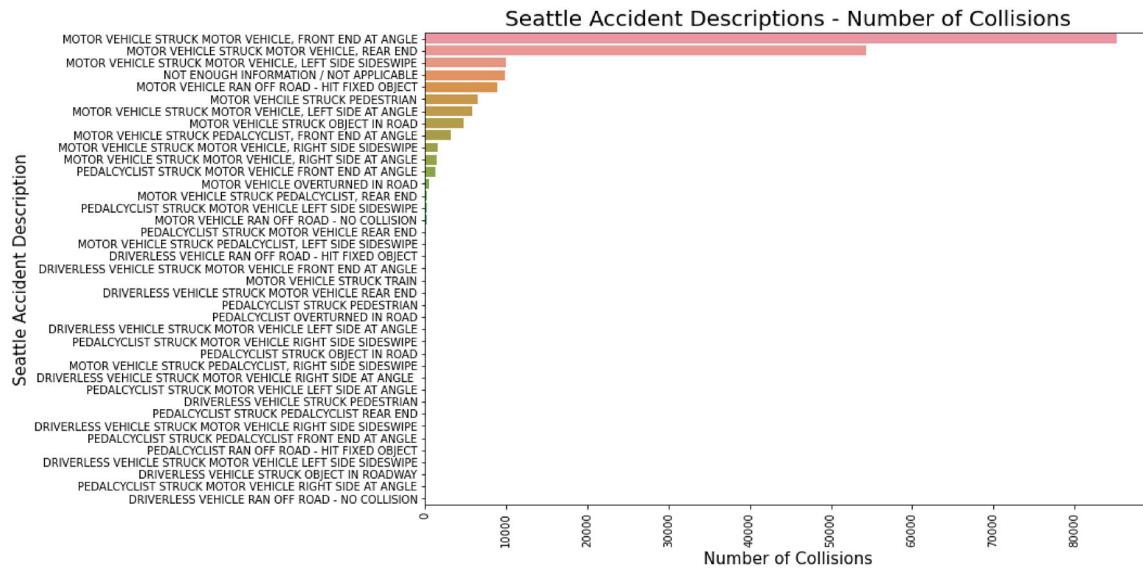
In [19]:

```
plot_barh(df['JUNCTIONTYPE'], 'Number of Collisions', 'Junction Type', 'Number of Collisions - Junction Type', rotate=90)
```



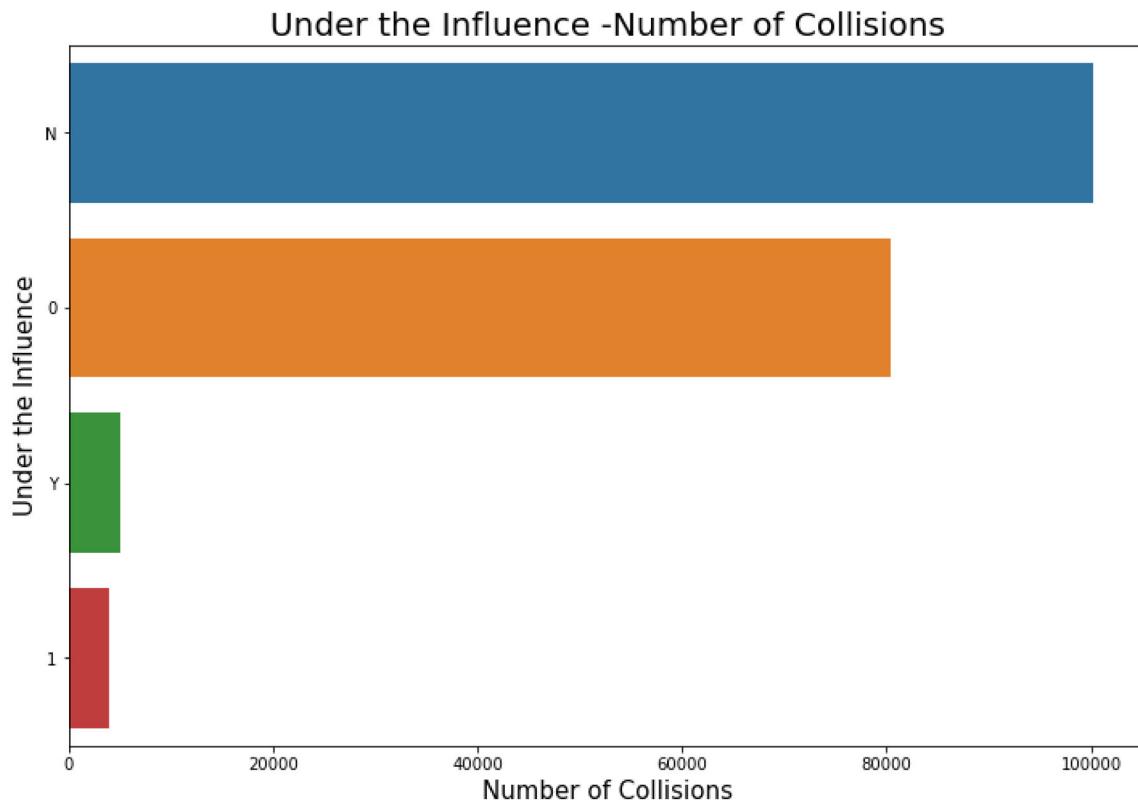
In [20]:

```
plot_barh(df['SDOT_COLDESC'], 'Number of Collisions', 'Seattle Accident Description', 'Seattle Accident Descriptions - Number of Collisions', rotate=90)
```



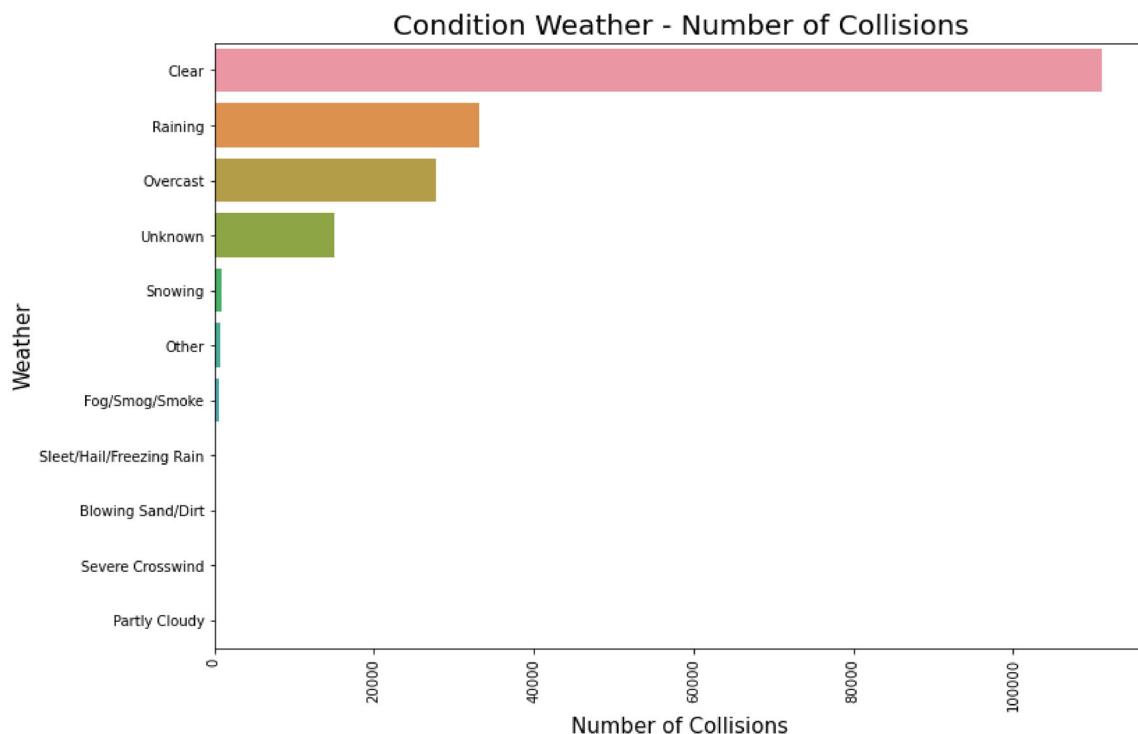
In [21]:

```
plot_barh(df['UNDERINFL'], 'Number of Collisions', 'Under the Influence', 'Under the Influence -Number of Collisions')
```



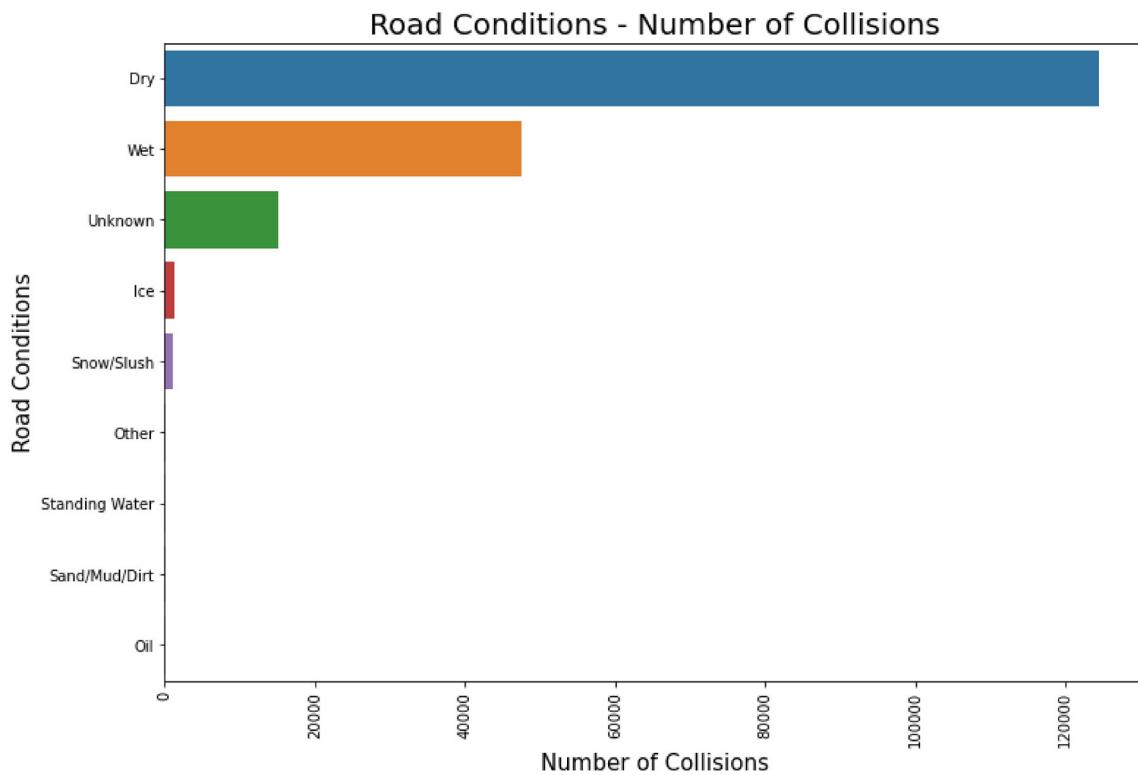
In [22]:

```
plot_barh(df['WEATHER'], 'Number of Collisions', 'Weather', 'Condition Weather - Number of Collisions', rotate=90)
```



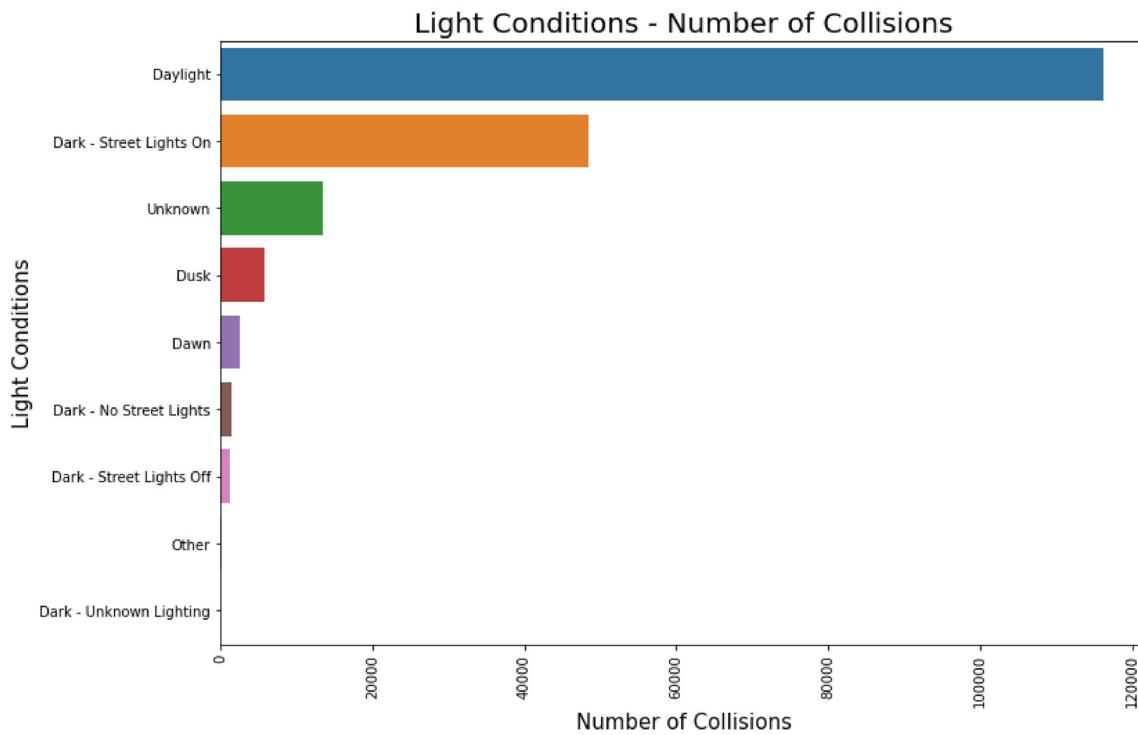
In [23]:

```
plot_barh(df['ROADCOND'], 'Number of Collisions', 'Road Conditions', 'Road Conditions - Number of Collisions', rotate=90)
```



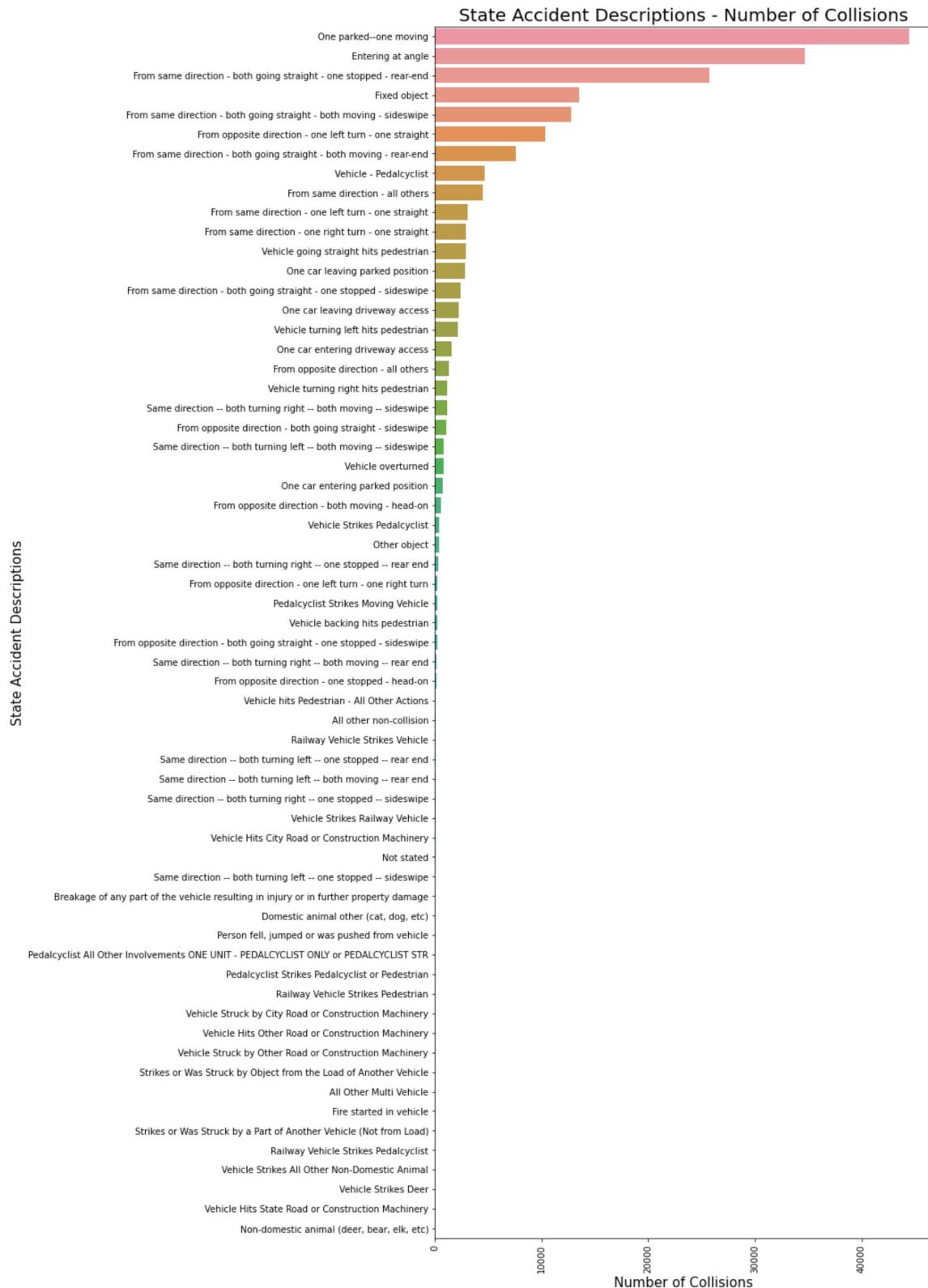
In [24]:

```
plot_barh(df['LIGHTCOND'], 'Number of Collisions', 'Light Conditions', 'Light Conditions - Number of Collisions', rotate=90)
```



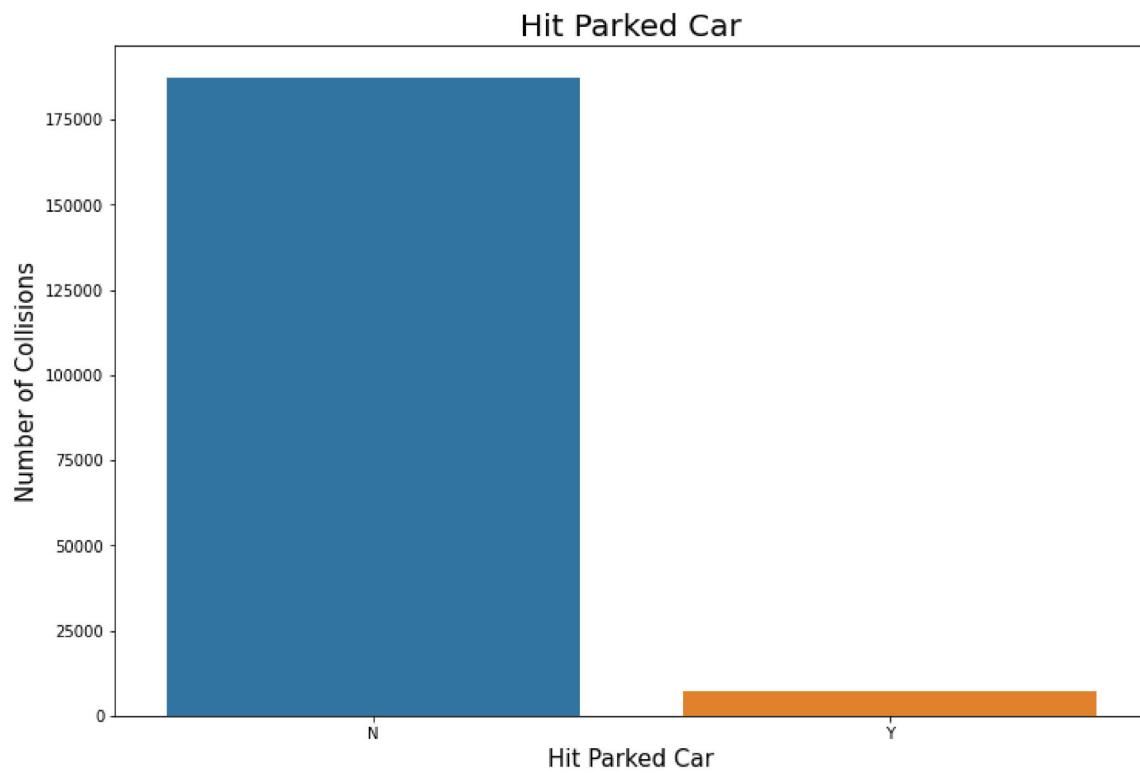
In [25]:

```
plot_barh(df['ST_COLDESC'], 'Number of Collisions', 'State Accident Descriptions', 'State Accident Descriptions - Number of Collisions', size=(10,25), rotate=90)
```



In [26]:

```
plot_bar(df[ 'HITPARKEDCAR' ], 'Hit Parked Car', 'Number of Collisions', 'Hit Parked Car'  
,)
```



## Analysis with speeding

In [107]:

```
car_velocity = df.loc[df['SPEEDING']=='Y']
car_velocity = car_velocity.dropna(axis=0, subset=['X', 'Y', 'SEVERITYCODE'])
print(car_velocity.shape[0])
car_velocity
```

8720

Out[107]:

SEVERITYCODE	X	Y	OBJECTID	INCKEY	COLDETKEY	REPORT
24	2	-122.279658	47.553405	33	1268	1268 3672
43	2	-122.337348	47.520472	53	56100	56100 2619
62	1	-122.376182	47.499490	74	32000	32000 1215
123	1	-122.333924	47.604678	140	29700	29700 1482
124	2	-122.322187	47.618733	141	1135	1135 3605
...	...	...	...	...	...	...
194385	1	-122.333532	47.690781	219206	308028	309308 3809
194414	1	-122.293204	47.542648	219238	309651	310931 3815
194428	2	-122.320008	47.625350	219255	309595	310875 E881
194481	2	-122.307643	47.541919	219317	308340	309620 3578

SEVERITYCODE	X	Y	OBJECTID	INCKEY	COLDETKEY	REPORT
--------------	---	---	----------	--------	-----------	--------

194549	1	-122.346793	47.662069	219399	308693	309973	3810
--------	---	-------------	-----------	--------	--------	--------	------

8720 rows × 38 columns

In [104]:

```
!conda install -c conda-forge folium=0.5.0 --yes
import folium

print('Folium installed and imported!')
```

Collecting package metadata (current\_repodata.json): ...working... done  
Solving environment: ...working... done

# All requested packages already installed.

Folium installed and imported!

In [108]:

```
latitude = 47.6062100
longitude = -122.3320700
# create map and display it
seattle_map = folium.Map(location=[latitude, longitude], zoom_start=12)

# display the map of seattle
```

In [115]:

```
# instantiate a feature group for the incidents in the dataframe
incidents = folium.map.FeatureGroup()

# Loop through the 8720 incidents feature group
for lat, lng, in zip(car_velocity.Y, car_velocity.X):
    incidents.add_child(
        folium.features.CircleMarker(
            [lat, lng],
            radius=1, # define how big you want the circle markers to be
            color='Red',
            fill=False,
        )
)

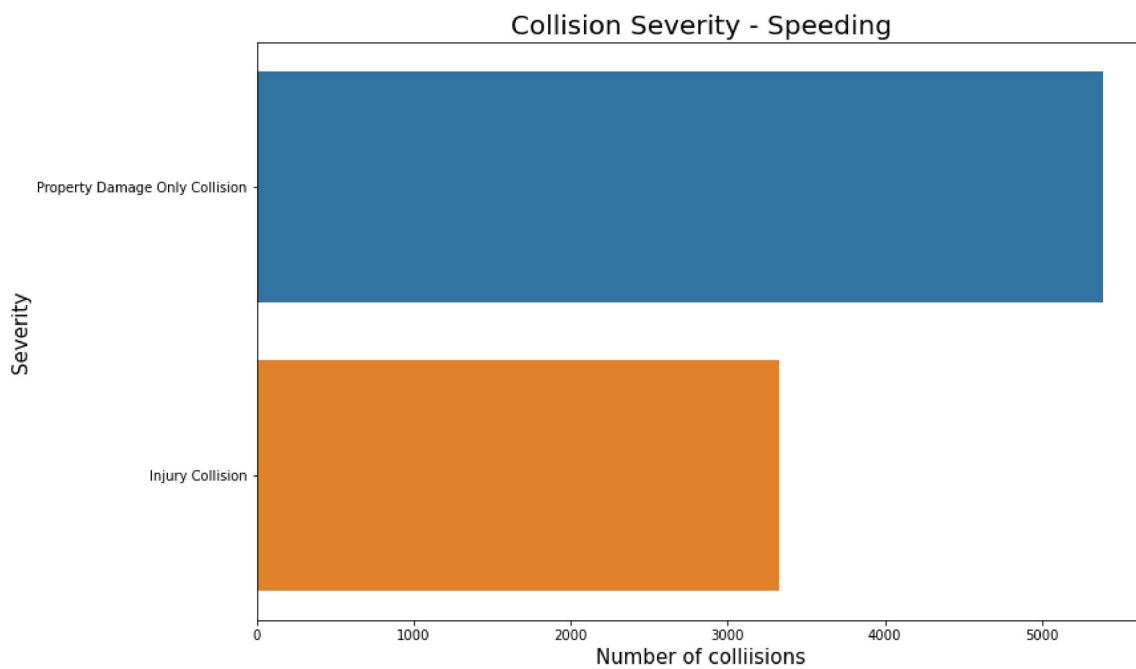
# add incidents to map
seattle_map.add_child(incidents)
```

Out[115]:

Make this Notebook Trusted to load map: File -> Trust Notebook

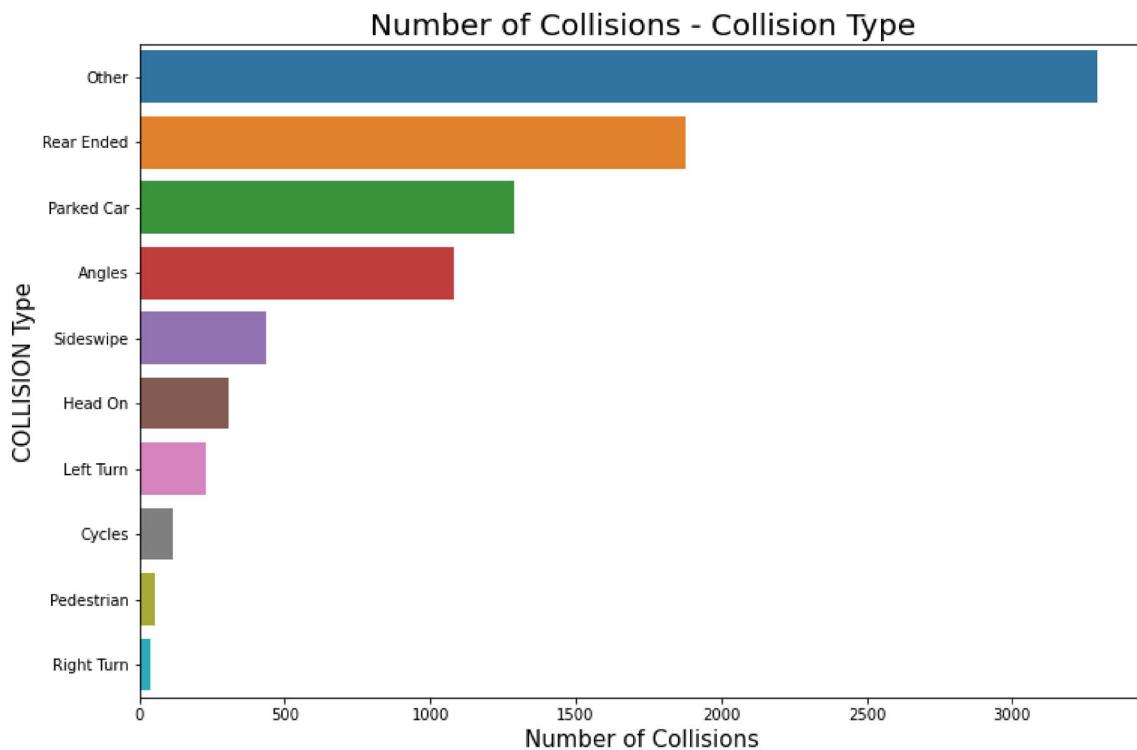
In [110]:

```
# Se analiza de los casos de velocidad el grado de daño  
plot_barh(car_velocity[ 'SEVERITYDESC' ],'Number of collisions','Severity','Collision Se  
verity - Speeding' )
```



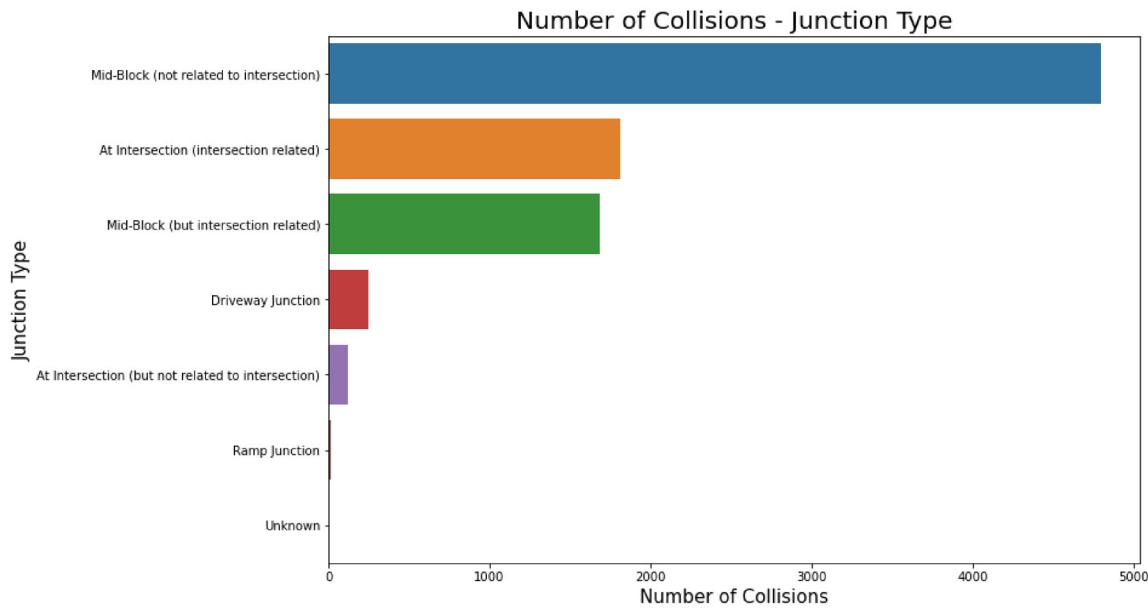
In [111]:

```
plot_barh(car_velocity['COLLISIONTYPE'], 'Number of Collisions', 'COLLISION Type', 'Number of Collisions - Collision Type')
```



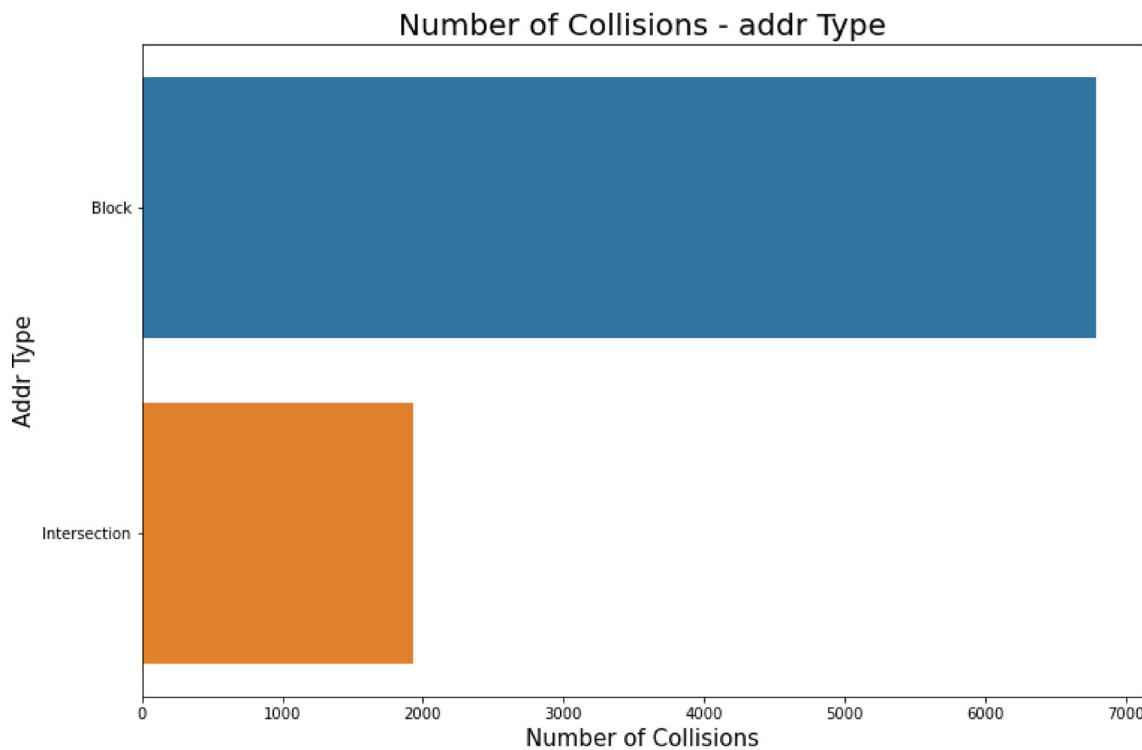
In [112]:

```
plot_barh(car_velocity['JUNCTIONTYPE'], 'Number of Collisions', 'Junction Type', 'Number of Collisions - Junction Type')
```



In [114]:

```
plot_barh(car_velocity['ADDRTYPE'], 'Number of Collisions', 'Addr Type', 'Number of Collisions - addr Type')
```



## Result and Discussion

## English

After conducting the first quick analysis on the data set that we obtained from a sample of accidents in Seattle, it is detailed.

- There are many more collisions that involve only property damage versus injuries.
- Collisions involving injuries tend to occur in and around the city center and major roads, while collisions without injuries are evenly distributed.
- Collisions occur more frequently in blocks than in intersections, very low accident rate in others.
- Most types of collisions are with parked cars. Rear angles and collisions are also common.
- The most frequent number of people involved in collisions is 2.
- Almost no pedestrians or bicyclists are involved in accidents.
- Most collisions involve 2 vehicles.
- Although the number of collisions has fluctuated over the past 16 years, it appears to be trending downward. The recession between 2007-2009 and the growth that followed may have had an impact.
- The number of collisions tends to increase slightly in October and decrease slightly in February.
- The number of collisions tends to go up slightly on Fridays and down slightly on Sundays.
- Most collisions do not involve alcohol or drugs.
- Most collisions occur when the weather is clear, the roads are dry, and it is daytime.
- The "parked car hit" function contradicts the "collision type" function.
- Speeding collisions occur mostly in the middle of the street and not at intersections.

## Español

Luego de realizar el primer análisis rápido sobre el conjunto de datos que obtuvimos de muestra de accidentes en Seattle se detalla.

- Hay muchas más colisiones que involucran solo daños a la property damage versus lesiones.
- Las colisiones que involucran lesiones tienden a ocurrir en y alrededor del centro de la ciudad y las carreteras principales, mientras que las colisiones sin lesiones se distribuyen uniformemente.
- Las colisiones ocurren con más frecuencia en los bloques que en las intersecciones, muy bajo tasa de accidentes en otras .
- La mayoría de los tipos de colisiones son con automóviles estacionados. Los ángulos y las colisiones traseras también son comunes.
- La cantidad más frecuente de personas involucradas en colisiones es 2.
- Casi no se involucra a peatones o ciclistas en los accidentes.
- La mayoría de las colisiones involucran a 2 vehículos.
- Aunque el número de colisiones ha fluctuado durante los últimos 16 años, parece tener una tendencia a la baja. La recesión entre 2007-2009 y el crecimiento que siguió después pueden haber tenido un impacto.
- El número de colisiones tiende a aumentar ligeramente en octubre y disminuir ligeramente en febrero.
- El número de colisiones tiende a subir ligeramente los viernes y bajar ligeramente los domingos.
- La mayoría de las colisiones no involucran alcohol o drogas.
- La mayoría de las colisiones ocurren cuando el clima está despejado, las carreteras están secas y es de día.
- La función de "coche aparcado atropellado" contradice la función de "tipo de colisión".
- Las colisiones por exceso de velocidad ocurren mayormente en la mitad de la calle y no en las intersecciones.

## Conclusion

## English

The data is consistent with each other. Many of the accidents occur under ideal driving conditions. This allows me to give my first opinion that the reasons for the accidents are due to carelessness or poor traffic infrastructure.

**As the first improvement proposal is to limit the maximum speeds in the central streets, since this would drastically reduce collisions.**

Seattle appears to be making progress toward its Vision Zero plan, but it has a long way to go. Some of the most relevant characteristics for predicting the severity of a collision can be the location, the type of direction and the type of collision. Additionally, since most days in Seattle are marked with cloudy skies, good weather and dry conditions can also be good characteristics for predicting the severity of the collision.

## Español

Los datos tiene coherencia entre si. Muchos de los accidentes ocurre en condiciones ideales para manejar. Esto me permite dar mi primera opinion sobre que los motivos de los accidentes se deben a descuidos o una mala infraestructura de tráfico.

**Como primera propuesta de mejora es limitar las velocidades maxima en las calles centricas, ya que asi se disminuiria drasticamente las coliciones.**

Seattle parece estar progresando hacia su plan Vision Zero, pero le queda mucho por hacer. Algunas de las características más relevantes para predecir la gravedad de una colisión pueden ser la ubicación, el tipo de dirección y el tipo de colisión. Además, dado que la mayoría de los días en Seattle están marcados con cielos nublados, el buen tiempo y las condiciones secas también pueden ser buenas características para predecir la gravedad de la colisión.