Arnav Tumbde
RCOEM Nagpur

Domain : Data Science

#Task 5

- Analyze traffic accident data to identify patterns related to road conditions, weather, and time of day. Visualize accident hotspots and contributing factors.

Screenshots of Source Code Jupyter Notebook :

In [7]:
```
import pandas as pd
```

In [6]:
```
data_filepath = "/content/drive/MyDrive/Kaggle/US_Accidents/US_Accidents_Dec20_Updated.csv"
```

In [8]:
```
df = pd.read_csv(data_filepath)
df.head(10)
```

Out[8]:

	ID	Severity	Start_Time	End_Time	Start_Lat	Start_Lng	End_Lat	End_Lng	Distance(mi)	Description	Number	Street	Side	City
0	A-1	2	2019-05-21 08:29:55	2019-05-21 09:29:40	34.808868	-82.269157	34.808868	-82.269157	0.000	Accident on Tanner Rd at Pennbrooke Ln.	439.0	Tanner Rd	R	Greenville
1	A-2	2	2019-10-07 17:43:09	2019-10-07 19:42:50	35.090080	-80.745560	35.090080	-80.745560	0.000	Accident on Houston Branch Rd at Providence Br...	3299.0	Providence Branch Ln	R	Charlotte
2	A-3	2	2020-12-13 21:53:00	2020-12-13 22:44:00	37.145730	-121.985052	37.165850	-121.988062	1.400	Stationary traffic on CA-17 from Summit Rd (CA...	NaN	Santa Cruz Hwy	R	Los Gatos
3	A-4	2	2018-04-17 16:51:23	2018-04-17 17:50:46	39.110390	-119.773781	39.110390	-119.773781	0.000	Accident on US-395 Southbound at Topsy Ln.	NaN	US Highway 395 S	R	Carson City
4	A-5	3	2016-08-31 17:40:49	2016-08-31 18:10:49	26.102942	-80.265091	26.102942	-80.265091	0.000	Accident on I-595 Westbound at Exit 4 / Pine I...	NaN	I-595 W	R	Fort Lauderdale
5	A-6	3	2018-10-17 16:40:36	2018-10-17 17:10:18	35.348240	-80.847221	35.348240	-80.847221	0.000	Three lanes blocked due to accident on I-77 No...	NaN	W W.T. Harris Blvd	R	Charlotte
6	A-7	4	2019-12-12 09:48:52	2019-12-12 10:18:05	39.523970	-107.777000	39.565780	-107.516950	14.153	Closed between CO-13/Taughenbaugh Blvd/Exit 90...	NaN	I-70 E	R	Rifle
7	A-8	2	2019-12-21 23:59:00	2019-12-22 00:32:06	34.034017	-118.026972	34.034017	-118.026972	0.000	At CA-60/Pomona Fwy - Accident.	NaN	CA-60 W	R	Whittier
8	A-9	2	2018-05-23 16:50:24	2018-05-23 22:50:24	35.863490	-86.831680	35.849480	-86.832530	0.969	At TN-248/Peytonsville Rd/Exit 61 - Accident. ...	425.0	Old Peytonsville Rd	R	Franklin
9	A-10	2	2019-01-30 08:44:18	2019-01-30 09:14:17	34.426330	-118.585100	34.420220	-118.581900	0.460	At Magic Mountain Pky - Accident. Hard shoulde...	NaN	Golden State Fwy S	R	Valencia

```
# Checking the columns in the data
df.columns
```

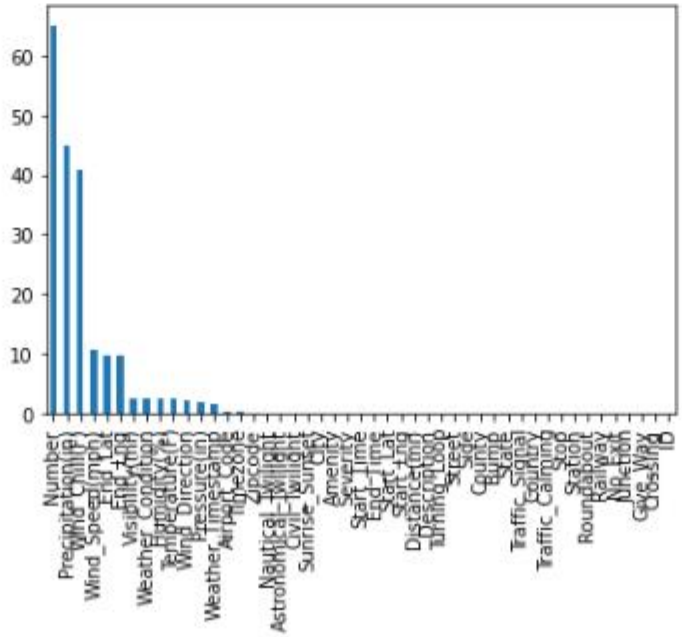
```
Index(['ID', 'Severity', 'Start_Time', 'End_Time', 'Start_Lat', 'Start_Lng',
       'End_Lat', 'End_Lng', 'Distance(mi)', 'Description', 'Number', 'Street',
       'Side', 'City', 'County', 'State', 'Zipcode', 'Country', 'Timezone',
       'Airport_Code', 'Weather_Timestamp', 'Temperature(F)', 'Wind_Chill(F)',
       'Humidity(%)', 'Pressure(in)', 'Visibility(mi)', 'Wind_Direction',
       'Wind_Speed(mph)', 'Precipitation(in)', 'Weather_Condition', 'Amenity',
       'Bump', 'Crossing', 'Give_Way', 'Junction', 'No_Exit', 'Railway',
       'Roundabout', 'Station', 'Stop', 'Traffic_Calming', 'Traffic_Signal',
       'Turning_Loop', 'Sunrise_Sunset', 'Civil_Twilight', 'Nautical_Twilight',
       'Astronomical_Twilight'],
      dtype='object')
```

```
df.describe()
```

	Severity	Start_Lat	Start_Lng	End_Lat	End_Lng	Distance(mi)	Number	Temperature(F)	Wind_Chill(F)	Humidity(%)	Pre
count	2.906610e+06	2.906610e+06	2.906610e+06	2.623789e+06	2.623789e+06	2.906610e+06	1.014938e+06	2.839386e+06	1.722751e+06	2.835340e+06	2.84
mean	2.288649e+00	3.653027e+01	-9.642676e+01	3.651733e+01	-9.620367e+01	3.980541e-01	6.789728e+03	6.098873e+01	5.499048e+01	6.537758e+01	2.96
std	5.541618e-01	5.013964e+00	1.775412e+01	5.016609e+00	1.765971e+01	1.592556e+00	1.697225e+04	1.845258e+01	2.219542e+01	2.287854e+01	9.09
min	1.000000e+00	2.455527e+01	-1.246238e+02	2.455527e+01	-1.246238e+02	0.000000e+00	0.000000e+00	-8.900000e+01	-8.900000e+01	1.000000e+00	0.00
25%	2.000000e+00	3.366453e+01	-1.178232e+02	3.364659e+01	-1.177020e+02	0.000000e+00	9.650000e+02	4.890000e+01	3.900000e+01	4.900000e+01	2.95
50%	2.000000e+00	3.609977e+01	-9.116690e+01	3.605898e+01	-9.105163e+01	0.000000e+00	3.093000e+03	6.300000e+01	5.800000e+01	6.800000e+01	2.99
75%	3.000000e+00	4.037505e+01	-8.085814e+01	4.033133e+01	-8.084679e+01	2.790000e-01	7.976000e+03	7.500000e+01	7.200000e+01	8.500000e+01	3.00
max	4.000000e+00	4.900220e+01	-6.711317e+01	4.907500e+01	-6.710924e+01	3.336300e+02	9.999997e+06	2.030000e+02	1.740000e+02	1.000000e+02	5.80

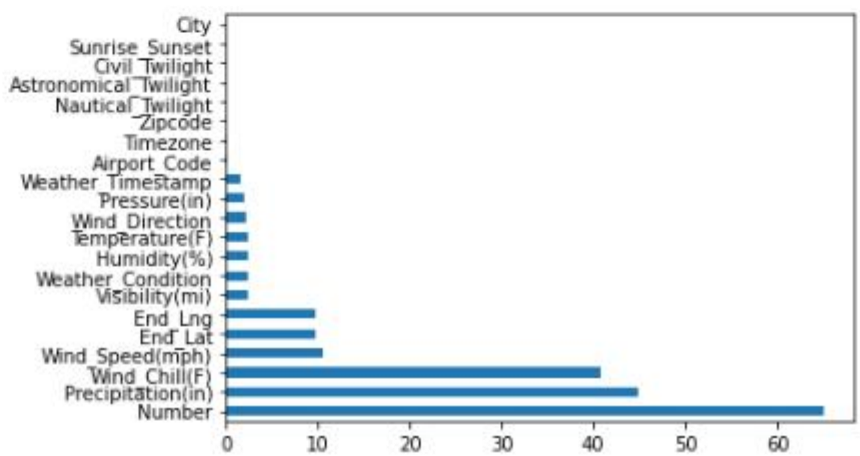
Screenshots of Visualizations:

Out[19]: <matplotlib.axes._subplots.AxesSubplot at 0x7fd7016fd910>

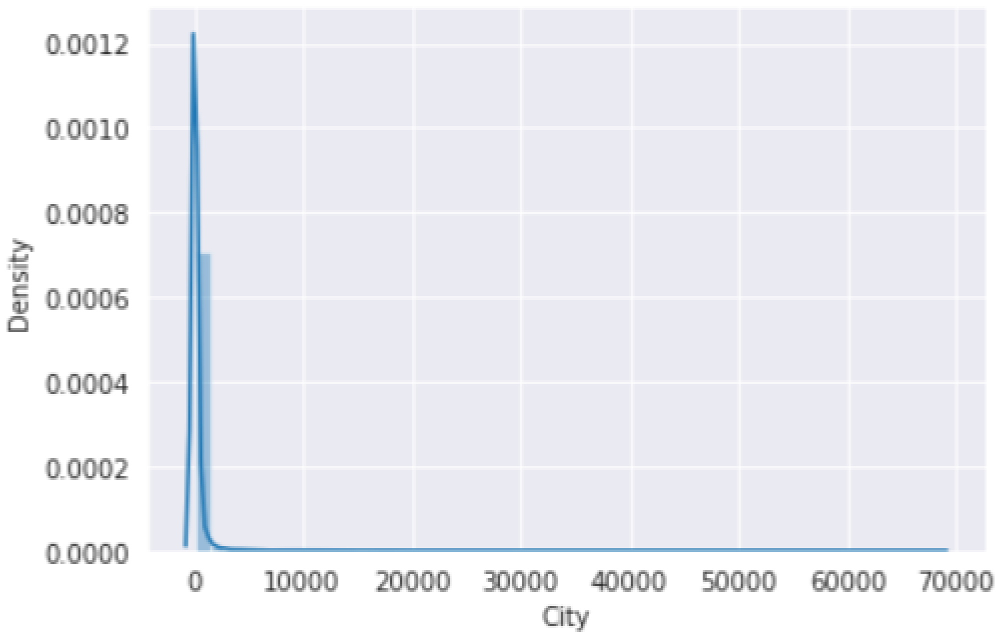


In [20]:
```
missing_data[missing_data!=0].plot(kind='barh')
```

Out[20]: <matplotlib.axes._subplots.AxesSubplot at 0x7fd701690750>



```
<matplotlib.axes._subplots.AxesSubplot at 0x7fd6f25f0fd0>
```



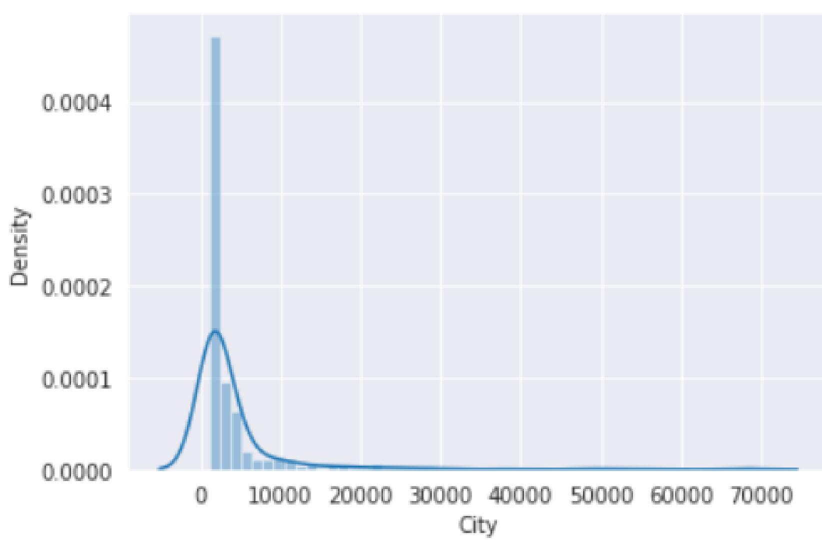
In [32]:

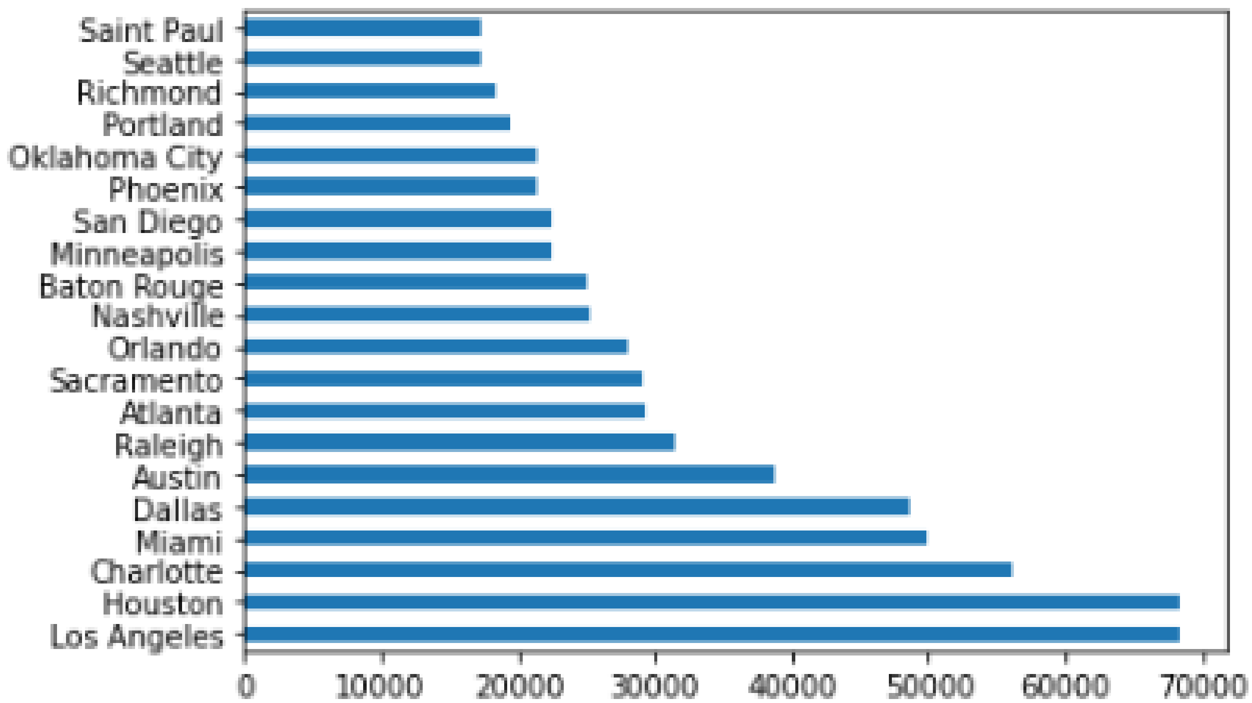
```
# Distribution of high accident cities
sns.distplot(high_accident_cities)
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.
ll be removed in a future version. Please adapt your code to
lity) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

Out[32]:

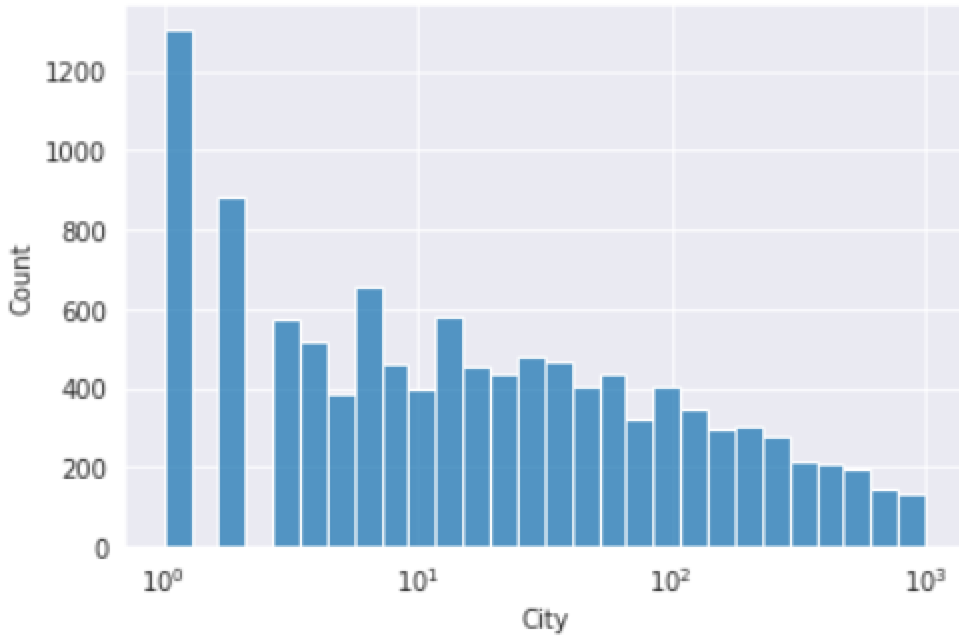
```
<matplotlib.axes._subplots.AxesSubplot at 0x7fd6e7c7e2d0>
```




Saint Paul
Seattle
Richmond
Portland
Oklahoma City
Phoenix
San Diego
Minneapolis
Baton Rouge
Nashville
Orlando
Sacramento
Atlanta
Raleigh
Austin
Dallas
Miami
Charlotte
Houston
Los Angeles
0
10000
20000
30000
40000
50000
60000
70000

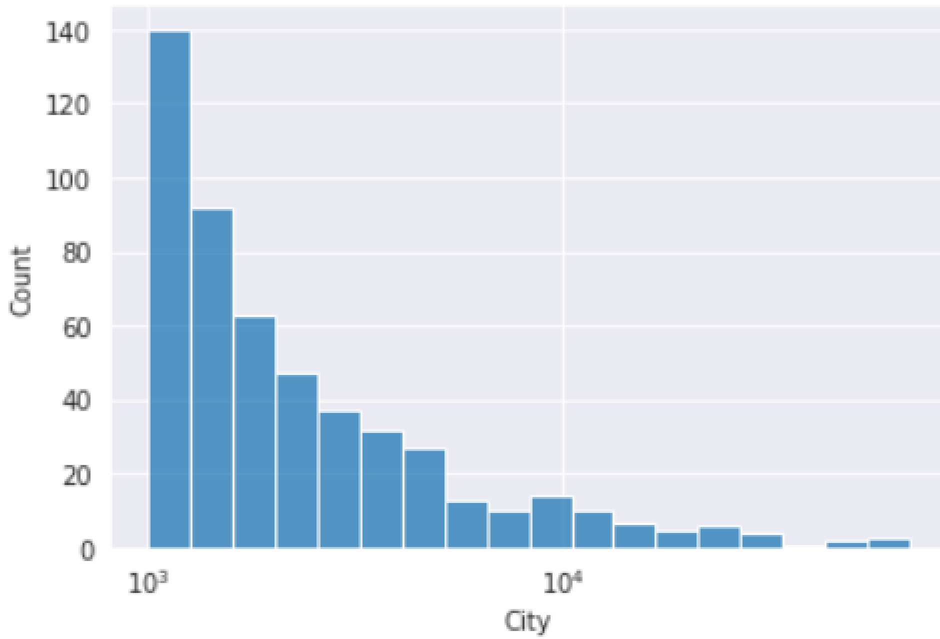
```
# Distribution of low accident cities
sns.histplot(low_accident_cities, log_scale=True)
```

<matplotlib.axes._subplots.AxesSubplot at 0x7fd6e7a3f110>



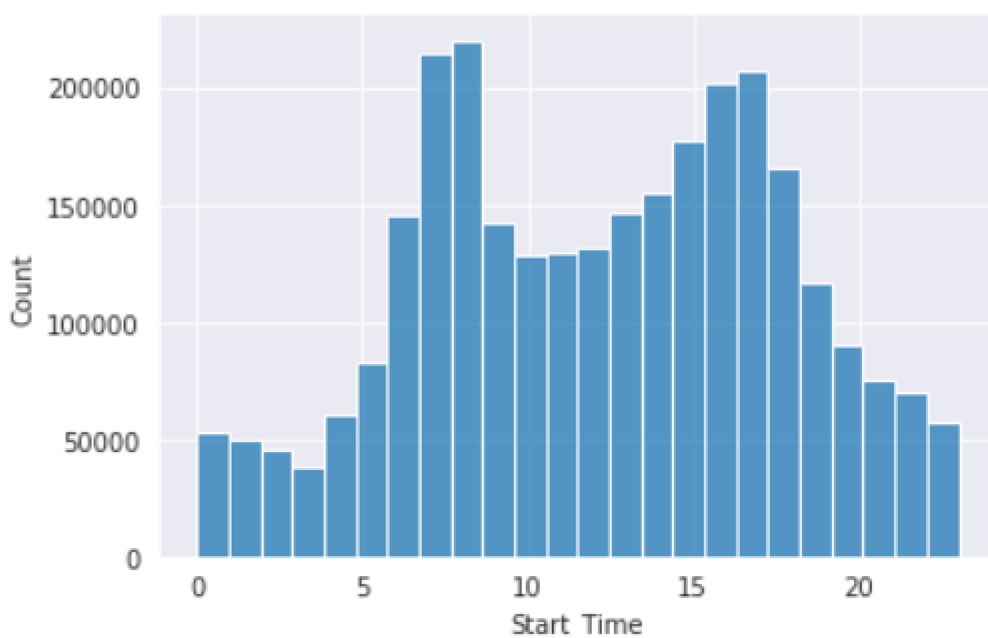
```python
# Distribution of high accident cities
sns.histplot(high_accident_cities, log_scale=True)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fd6e7a57650>
```



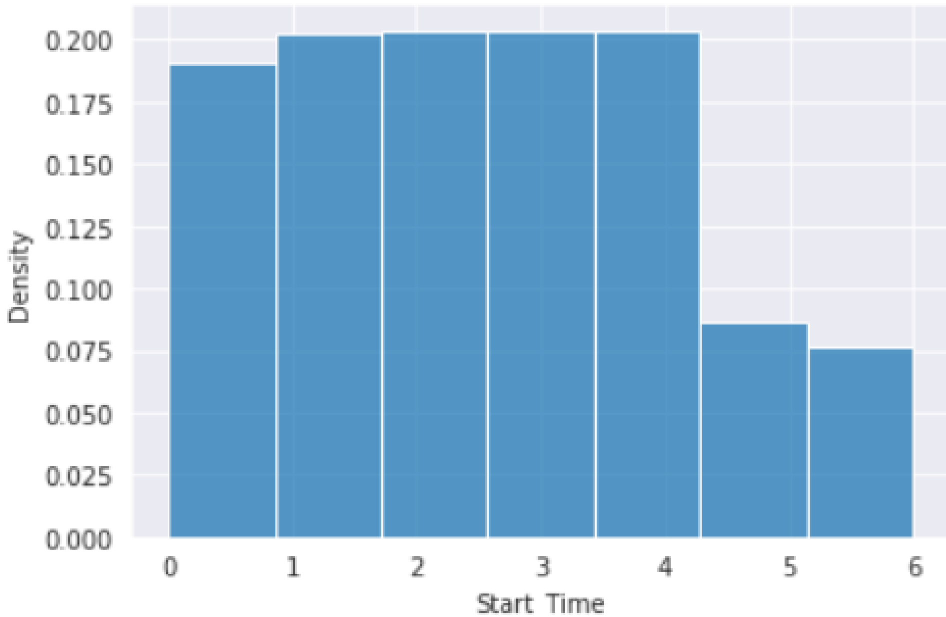

```python
sns.histplot(df.Start_Time.dt.hour, bins=24)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fd6e7b41090>
```



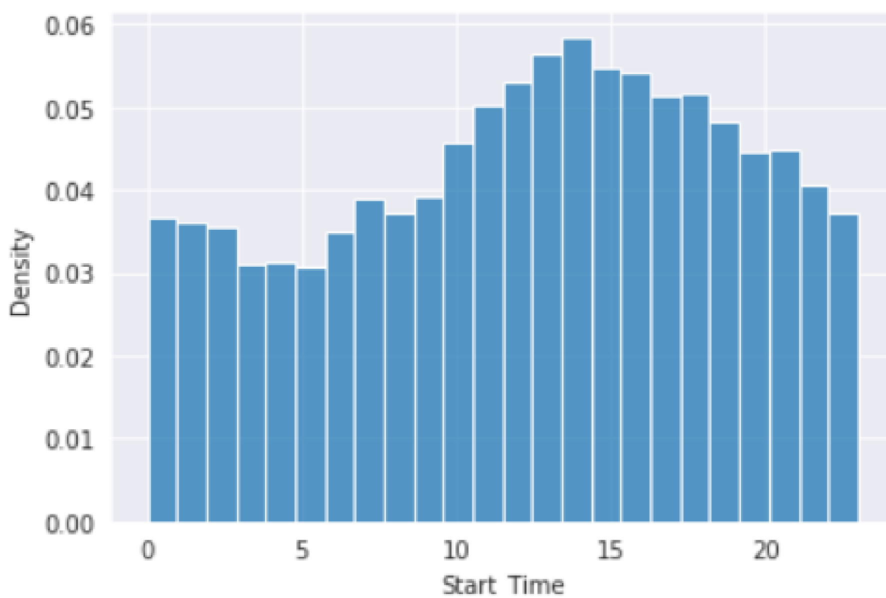
```python
sns.histplot(df.Start_Time.dt.dayofweek, bins=7, stat='density')
```

<matplotlib.axes._subplots.AxesSubplot at 0x7fd6e76aea50>

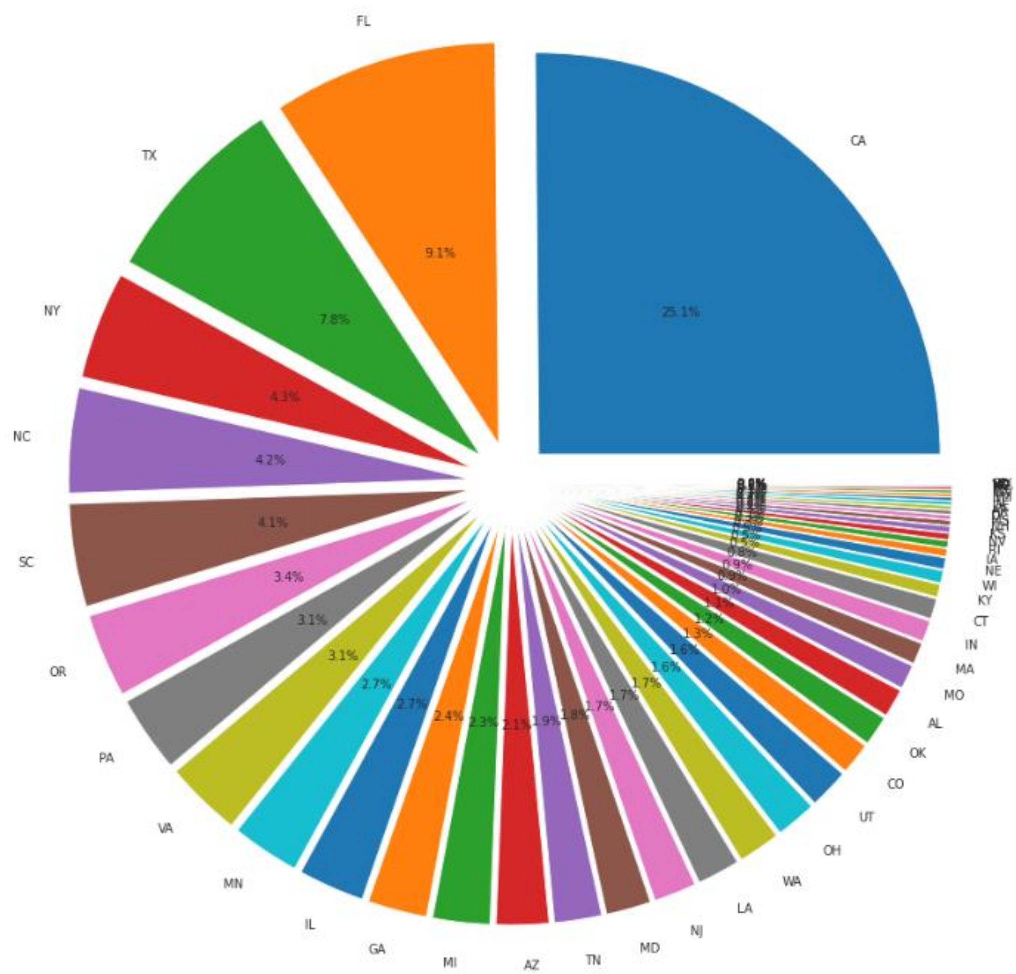



```python
sns.histplot(sundays_start_time.dt.hour, bins=24, stat='density')
```

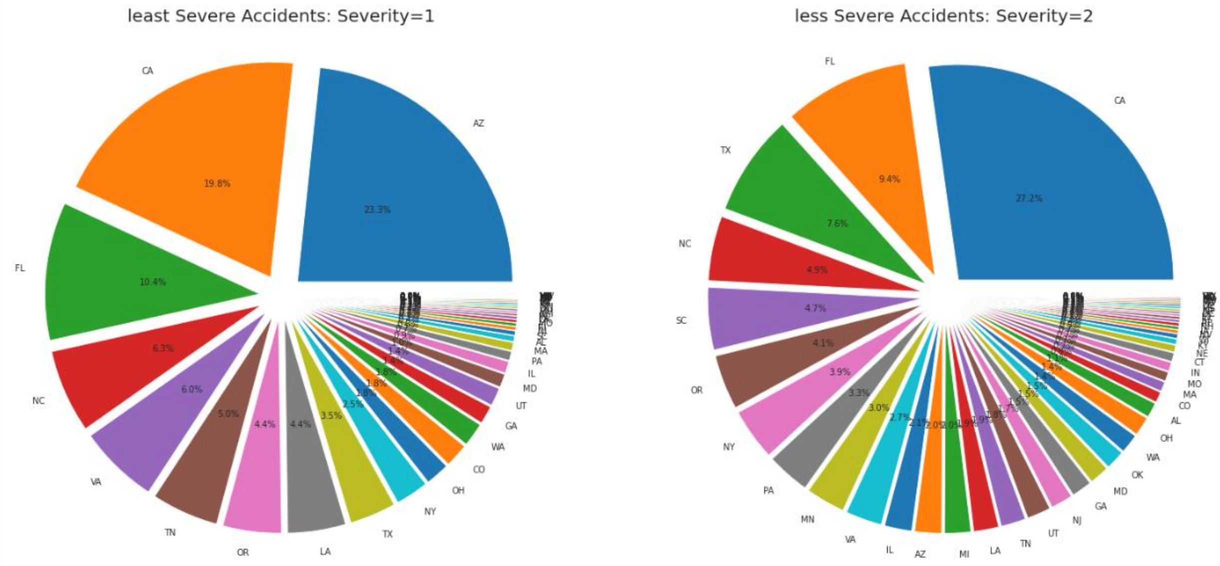
<matplotlib.axes._subplots.AxesSubplot at 0x7fd6e75ce750>



```python
pie, ax = plt.subplots(figsize=[15,15])
labels = df.State.value_counts().keys()
plt.pie(x=df.State.value_counts(), autopct="%.1f%%", explode=[0.1]*len(df.State.value_counts()), labels=labels, pctdistance=0.5)
plt.show();
```

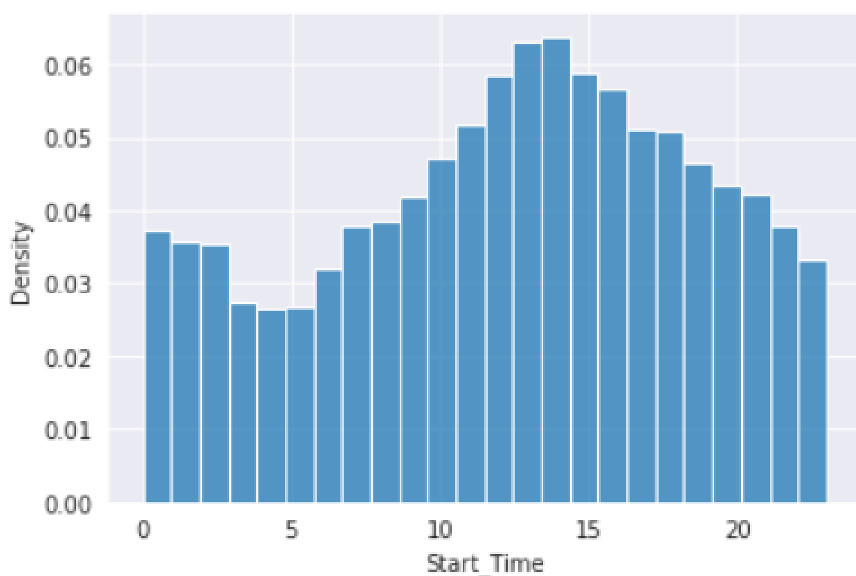


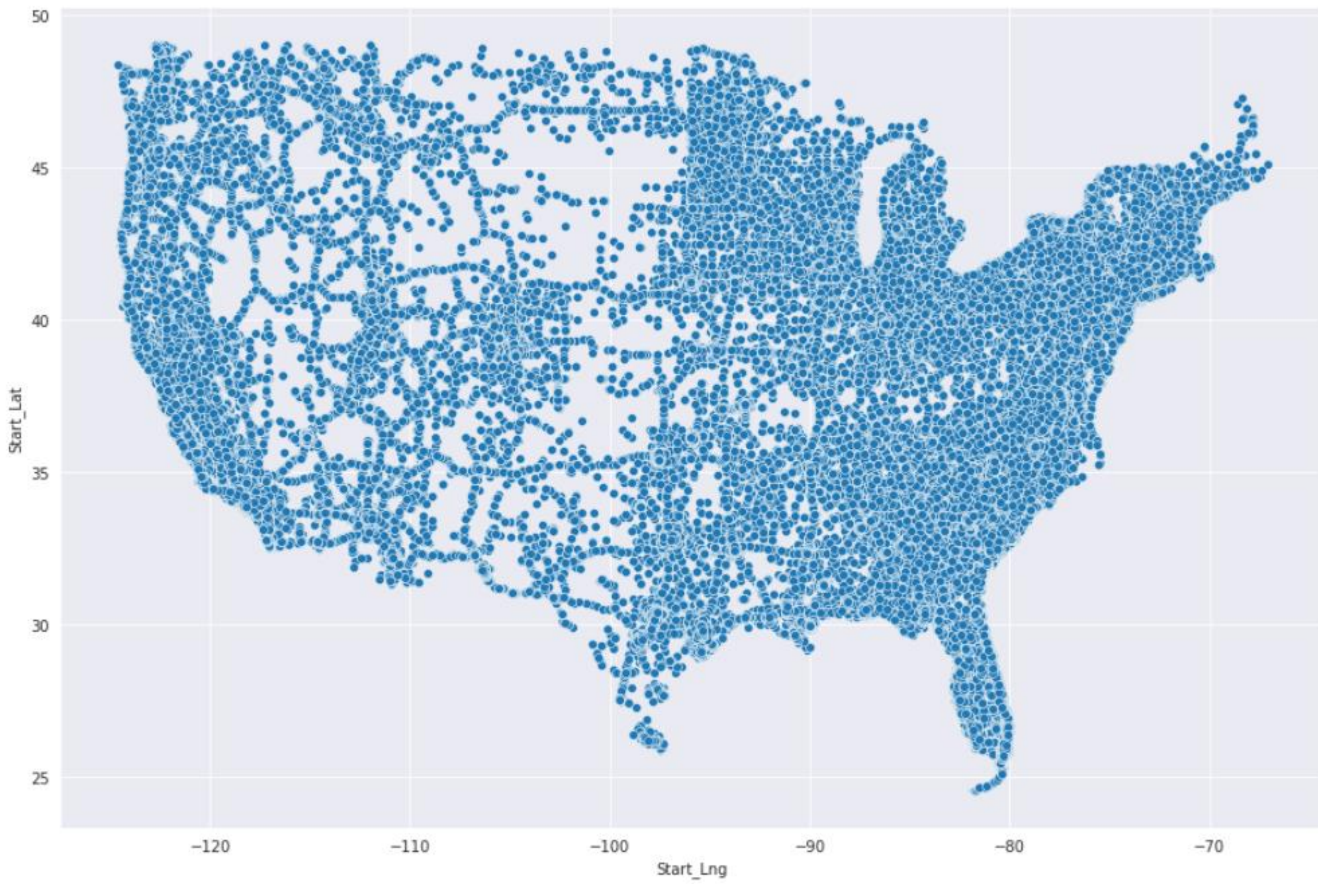
```
Text(0.5, 1.0, 'Most Severe Accidents: Severity=4')
```




```
saturdays_start_time = df.Start_Time[df.Start_Time.dt.dayofweek == 5]
sns.histplot(saturdays_start_time.dt.hour, bins=24, stat='density')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fd6e74f8190>
```





a. About the Dataset

For this task, I utilized the US Accidents dataset, which contains detailed information about traffic accidents across the United States from 2016 to 2023. The dataset includes various attributes that describe the accident characteristics, road conditions, weather conditions, time of day, and geographical location.

b. Explanation of the Concept

The main goal of this task was to analyze traffic accident data to identify patterns related to road conditions, weather conditions, and time of day. By exploring these patterns, we aimed to understand where and when accidents are more likely to occur and what factors contribute to these accidents.

c. Outcome of the Analysis

1. **Accident Hotspots**: I created a heatmap to visualize accident hotspots across different regions in the United States. This helped identify areas with high accident frequencies, highlighting potential areas for increased safety measures.
2. **Road Conditions and Accidents**: A bar chart was used to show how different road conditions (e.g., dry, wet, icy) correlate with the number of accidents. This analysis provided insights into which road conditions are most hazardous.
3. **Weather Conditions and Accidents**: A line chart or bar chart was utilized to explore how weather conditions (e.g., clear, rainy, foggy) impact accident rates. This visualization helped identify weather-related accident patterns.
4. **Time of Day Analysis**: A time-series analysis or bar chart showed the distribution of accidents throughout the day. This visualization revealed peak accident times and insights into how accident rates vary by time of day.

d. Conclusion

- **Accident Hotspots**: The heatmap highlighted regions with higher accident frequencies, guiding efforts for targeted safety improvements and traffic management strategies.
- **Road and Weather Conditions**: Analyzing road and weather conditions provided insights into factors contributing to accidents, enabling proactive measures like road maintenance and weather alerts.
- **Temporal Analysis**: Understanding accident patterns throughout the day helps in scheduling traffic enforcement and emergency response resources effectively.

By analyzing and visualizing these patterns in traffic accident data, we gain valuable insights that can inform policies, infrastructure improvements, and public safety initiatives to reduce accident rates and enhance road safety across the United States.