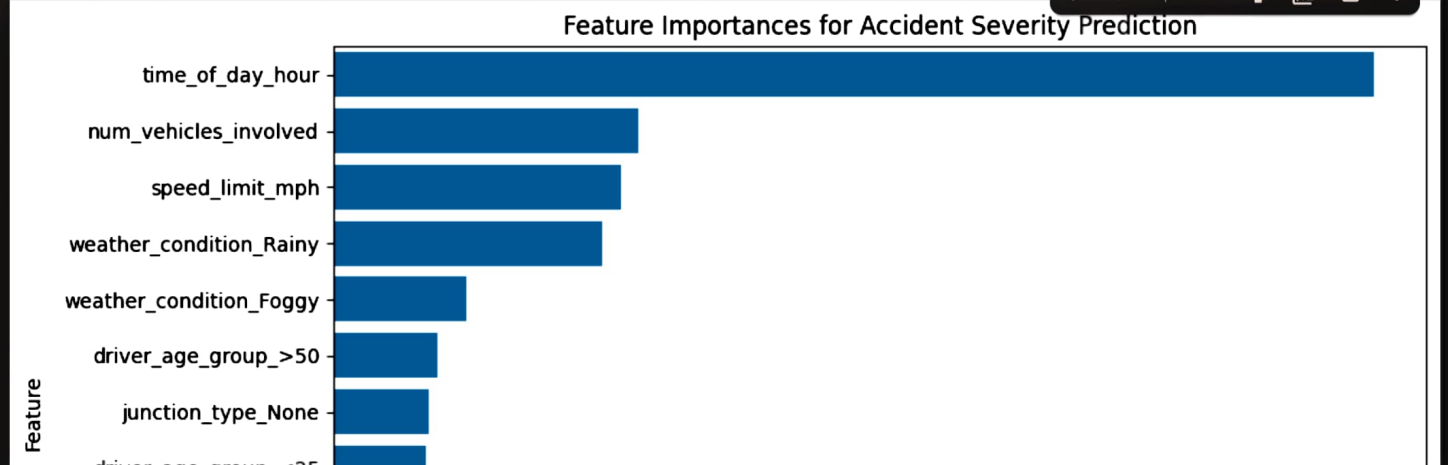
```
--- Generating Simulated Data ---
Generated 5000 records.
   accident_id       date  time_of_day_hour     road_type weather_condition  \
0            0 2023-04-13                23       Highway             Clear
1            1 2023-12-15                 9    Rural Road             Rainy
2            2 2023-09-28                 6  Urban Street             Rainy
3            3 2023-04-17                16       Highway             Foggy
4            4 2023-03-13                 3    Rural Road             Clear

   speed_limit_mph  num_vehicles_involved junction_type driver_age_group  \
0               30                      4    Crossroads             <25
1               75                      4    Roundabout           25-50
2               45                      2    Roundabout             >50
3               45                      2          None           25-50
4               45                      1          None           25-50

  severity
0  Serious
1  Serious
```

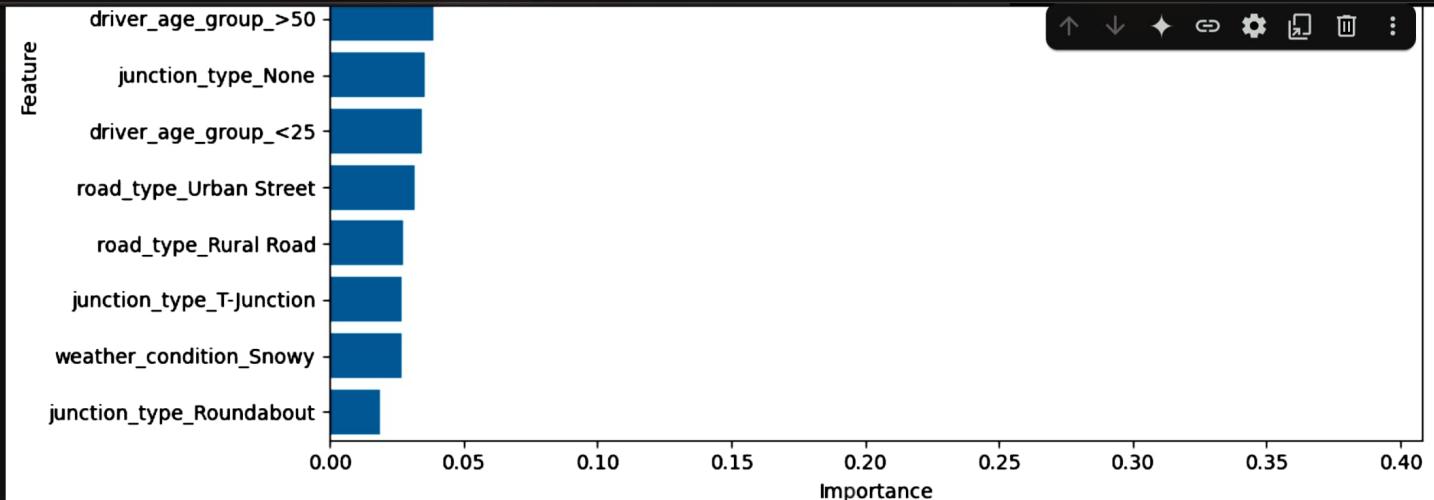
```
  severity
0  Serious
1  Serious
2  Serious
3  Serious
4    Minor

--- Data Info ---
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5000 entries, 0 to 4999
Data columns (total 10 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   accident_id           5000 non-null   int64
 1   date                  5000 non-null   datetime64[ns]
 2   time_of_day_hour      5000 non-null   int64
 3   road_type             5000 non-null   object
 4   weather_condition     5000 non-null   object
 5   speed_limit_mph       5000 non-null   int64
 6   num_vehicles_involved 5000 non-null   int64
```

```
 6   num_vehicles_involved  5000 non-null   int64
 7   junction_type          5000 non-null   object
 8   driver_age_group       5000 non-null   object
 9   severity               5000 non-null   object
dtypes: datetime64[ns](1), int64(4), object(5)
memory usage: 390.8+ KB

--- Data Preprocessing ---
Features after encoding: ['time_of_day_hour', 'speed_limit_mph', 'num_vehicles_involved', 'road_type_Rural Road', 'roa
Shape of features (X): (5000, 13)
Shape of target (y): (5000,)
Training set size: 4000 samples
Testing set size: 1000 samples

--- Model Training (Random Forest Classifier) ---
Model training complete.

--- Model Prediction and Evaluation ---
Accuracy: 0.5460
```

```
Accuracy: 0.5460

Classification Report:
              precision    recall  f1-score   support

       Minor       0.60      0.64      0.62       331
     Serious       0.59      0.63      0.61       484
       Fatal       0.22      0.15      0.18       185

    accuracy                           0.55      1000
   macro avg       0.47      0.48      0.47      1000
weighted avg       0.52      0.55      0.53      1000


Confusion Matrix:
 [[213  93  25]
 [107 305  72]
 [ 37 120  28]]

# Basic Analysis & Visualization
```

```
--- Basic Analysis & Visualization ---
```





```
<ipython-input-1-5807591ab518>:113: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hu

  sns.countplot(x='severity', data=df, palette='viridis', order=['Minor', 'Serious', 'Fatal'])
```