Predicting Injuries at Theme Parks

By Matthew Coulombe

Problem Statement

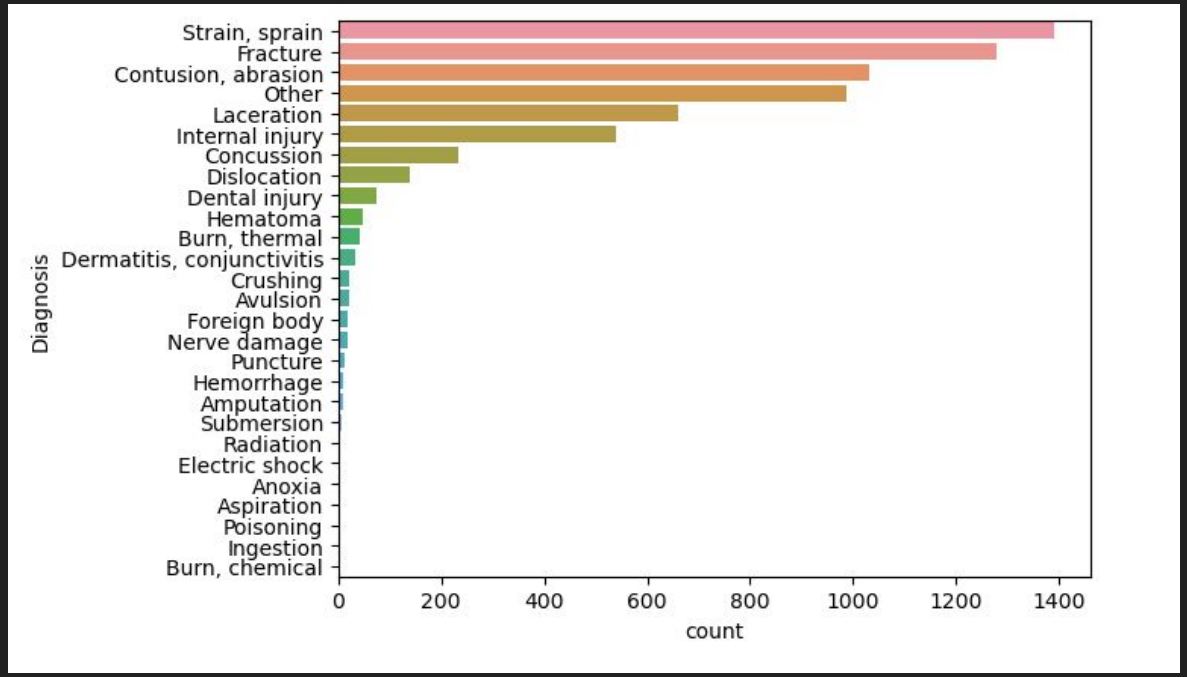
How can we predict a common injury diagnosis given characteristics of a rider and ride?

The Data

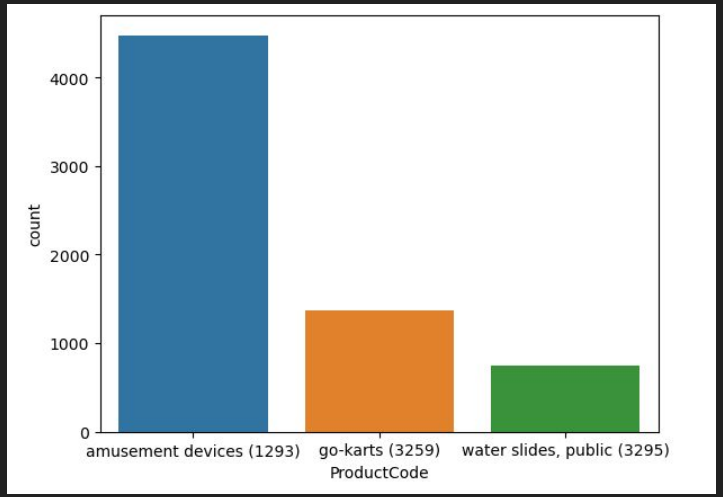
- NEISS hospital data from 2013-2017
  - https://ridesdatabase.org/saferparks/data/
- Target Value: Diagnosis

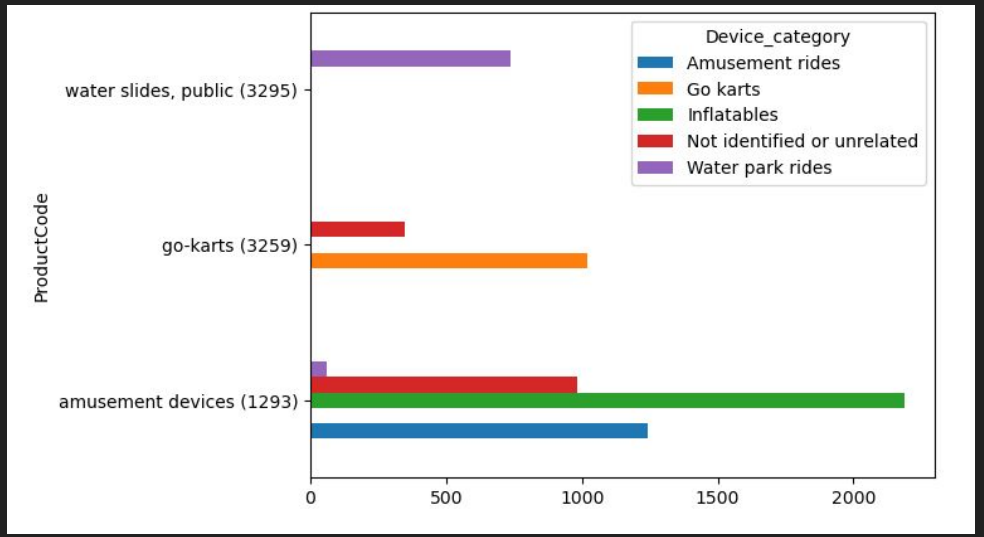
Excel/Tab Variable	SAS Variable	Description/Code book
CPSC_Case_Number	NEK	CPSC case number
Treatment_Date	Trmt_date	Date of Treatment
Age	Age	Age of Patient
Sex	Sex	Sex of Patient
Race	Race	Race of Patient
Other_Race	Raceoth	Description of Other Race (Used with Race=3)
Body_Part	Bdpt	Injured Body Part
Diagnosis	Diag	Injury Diagnosis
Other_Diagnosis	Diagoth	Description of Other Diagnosis (Used with Diag...
Disposition	Disp	Disposition
Location	Loc	Incident Location
Product_1	Prod1	Product Code (See NEISS Coding Manual for Deta...
Narrative	Narr	Description of Injury Event
Stratum	Stratum	Design Variable-Stratum
PSU	PSU	Design Variable-Primary Sampling Unit (PSU)
Weight	Wt	Statistical Weight for National Estimates

Exploratory Data Analysis - Diagnosis

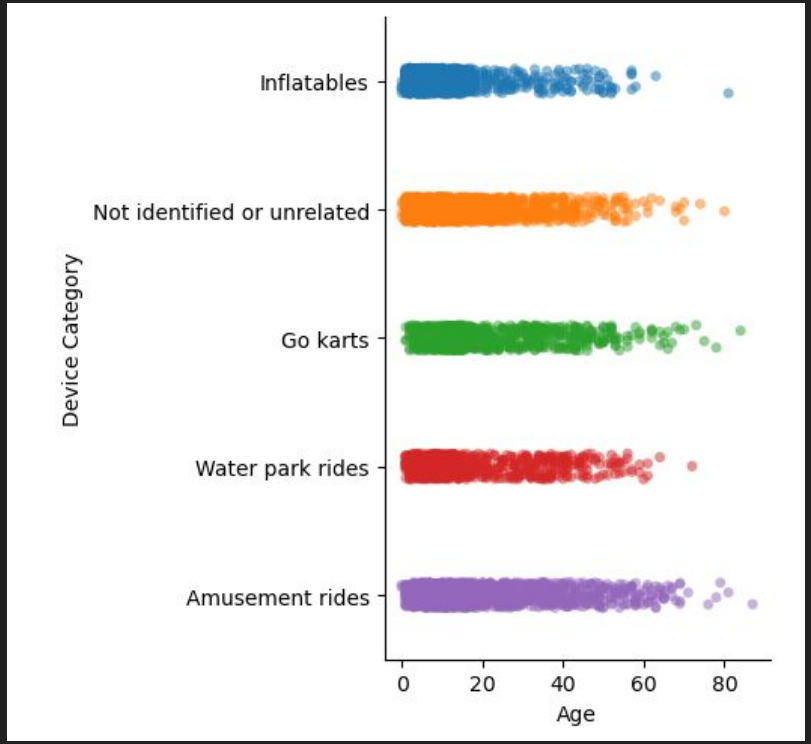


Exploratory Data Analysis - Product Codes

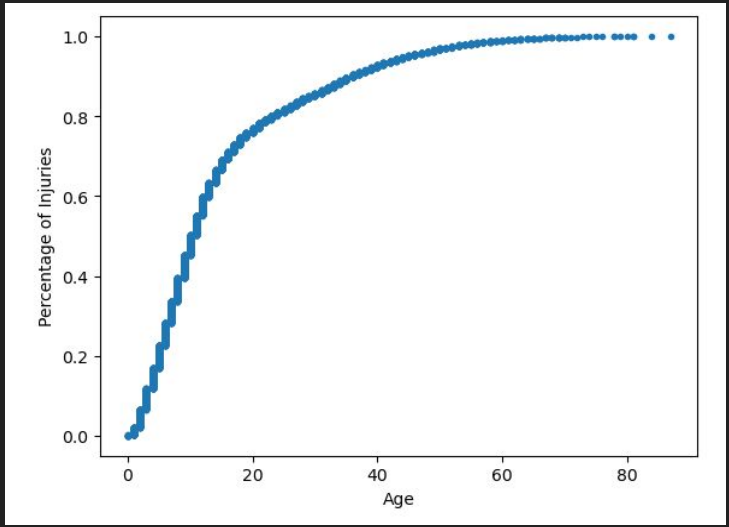


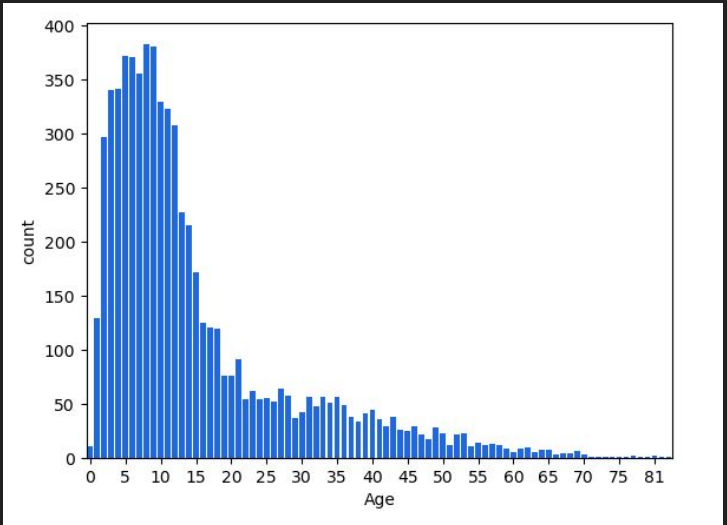


Exploratory Data Analysis - Device Category

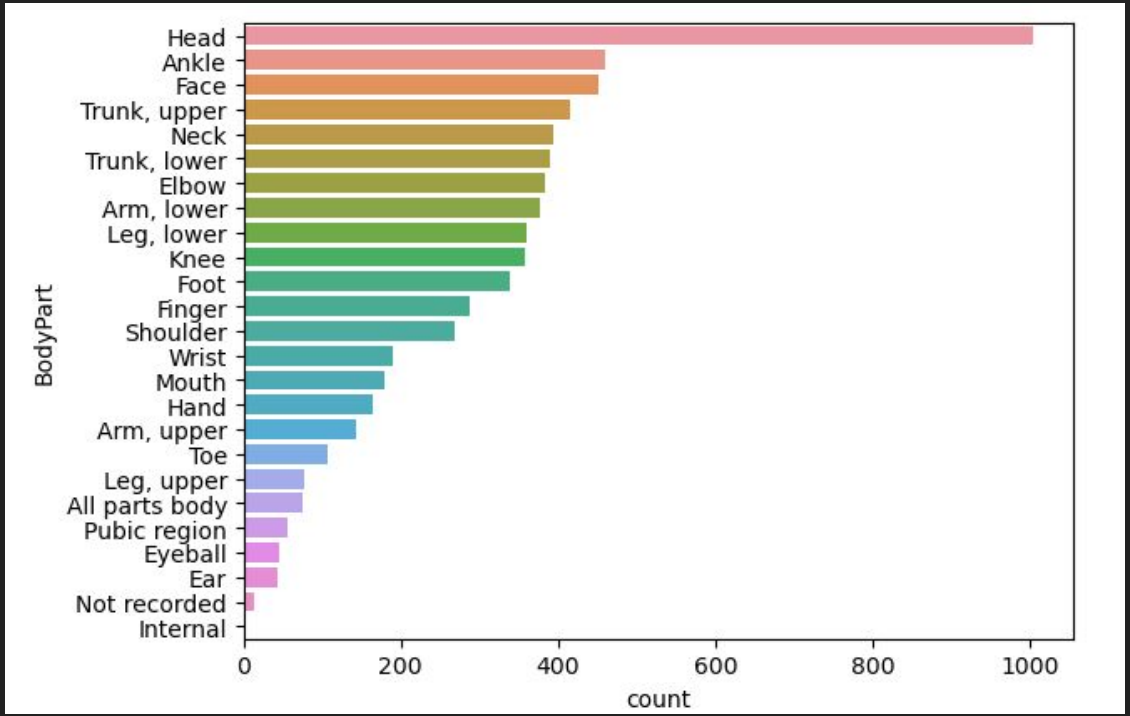


Exploratory Data Analysis - Age

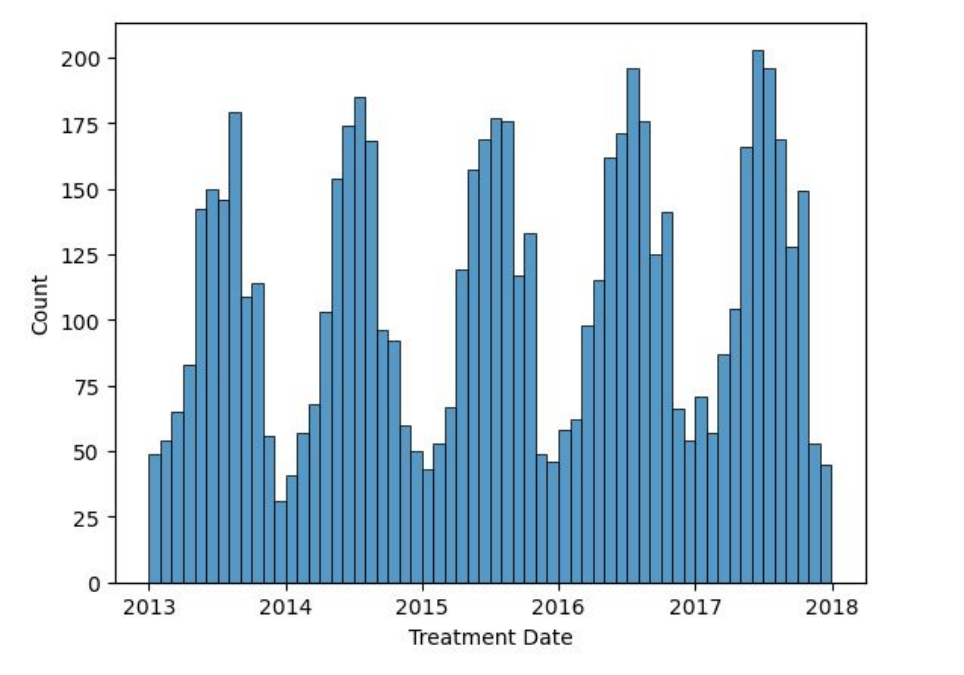




Exploratory Data Analysis - Body Part



Exploratory Data Analysis - Treatment Date



Preprocessing

CPSC_Case_Number	Diagnosis	Split_Diagnosis
180125260	Fracture	common
180108428	Dental injury	uncommon
180120413	Other	uncommon
180125238	Fracture	common
180135290	Strain, sprain	common
...	...	...
130113361	Contusion, abrasion	common
130109590	Laceration	uncommon
130113339	Nerve damage	uncommon
130109054	Hematoma	uncommon
130123446	Dental injury	uncommon

CPSC_Case_Number	day_of_week	Treatment_Date	day_of_week_S
180125260	6	2017-12-31	Sunday
180108428	6	2017-12-31	Sunday
180120413	6	2017-12-31	Sunday
180125238	5	2017-12-30	Saturday
180135290	5	2017-12-30	Saturday
...	...	...	...
130113361	3	2013-01-03	Thursday
130109590	2	2013-01-02	Wednesday
130113339	2	2013-01-02	Wednesday
130109054	1	2013-01-01	Tuesday
130123446	1	2013-01-01	Tuesday

Models

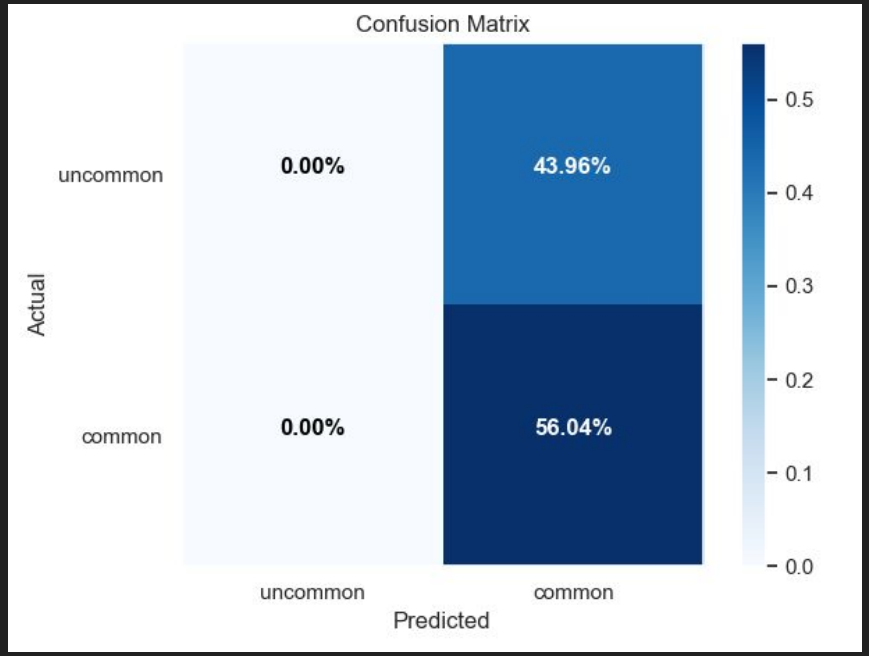
Three Models:

- Baseline Model
- Logistic Regression Model
- Random Forest Model

5 Scoring Metrics:

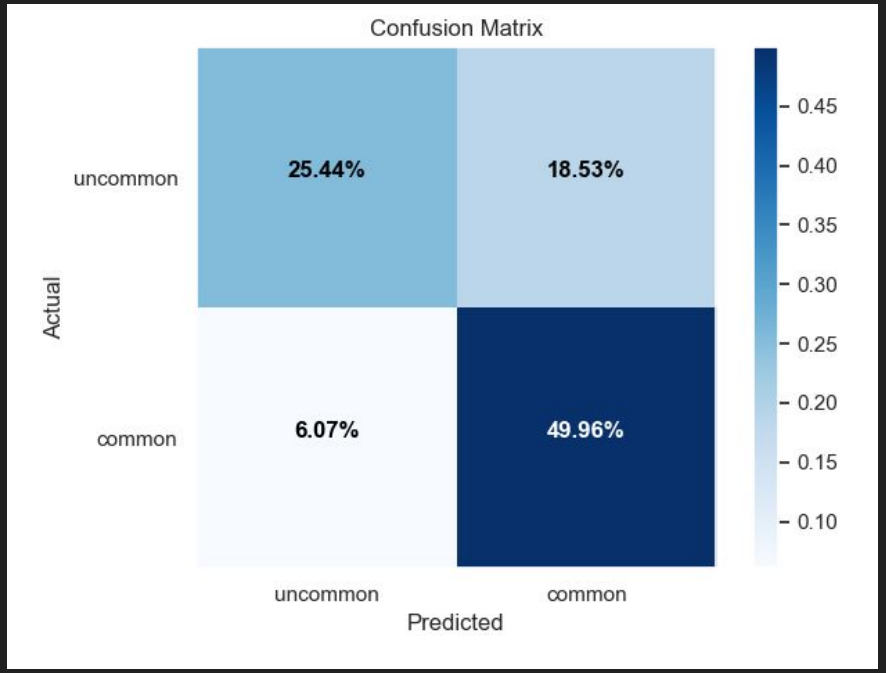
- Accuracy
- Precision
- Recall
- F1
- Balanced Accuracy

Models - Baseline



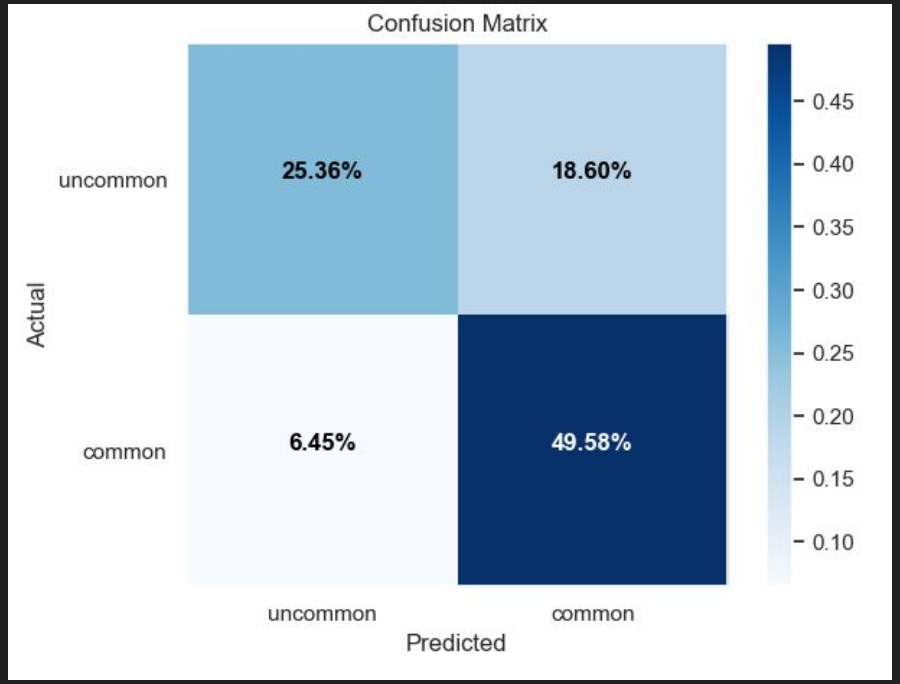

```
Accuracy: 0.56
Precision: 0.56
Recall: 1.0
F1: 0.718
Balanced Accuracy: 0.5
```

Models - Logistic Regression




```
Accuracy: 0.754
Precision: 0.729
Recall: 0.892
F1: 0.802
Balanced Accuracy: 0.735
```

Models - Random Forest




```
Accuracy: 0.749
Precision: 0.727
Recall: 0.885
F1: 0.798
Balanced Accuracy: 0.731
```

Conclusion

- Most common ride type
  - Inflatables
- Most commonly injured part of the body
  - The head
- Most common injury Diagnosis
  - Strain, sprain
  - Fracture
  - Contusion, abrasion
- Best Model
  - Logistic Regression

Next Steps

- Determine policies and procedures to prevent ride injury
- Tune Hyperparameters
- Breakdown Common/Uncommon injuries
- Try other models for better performance
- Dive deeper into the results of our model

Limitations

- Very basic modeling procedure
- PSU, Weight, and Stratum (Fields from Dataset)
- Data is for the type of injury, assuming there is an incident, not the rate at which injuries occur.