DISNEY WORLD'S BEST RIDES

Machine Learning Problem

- What features of a ride or attraction affect the star rating?
  - AKA what makes a ride “*great*” rather than good

- Benefits Disney Parks and their subsidiaries
  - Development of future attractions

Introduction

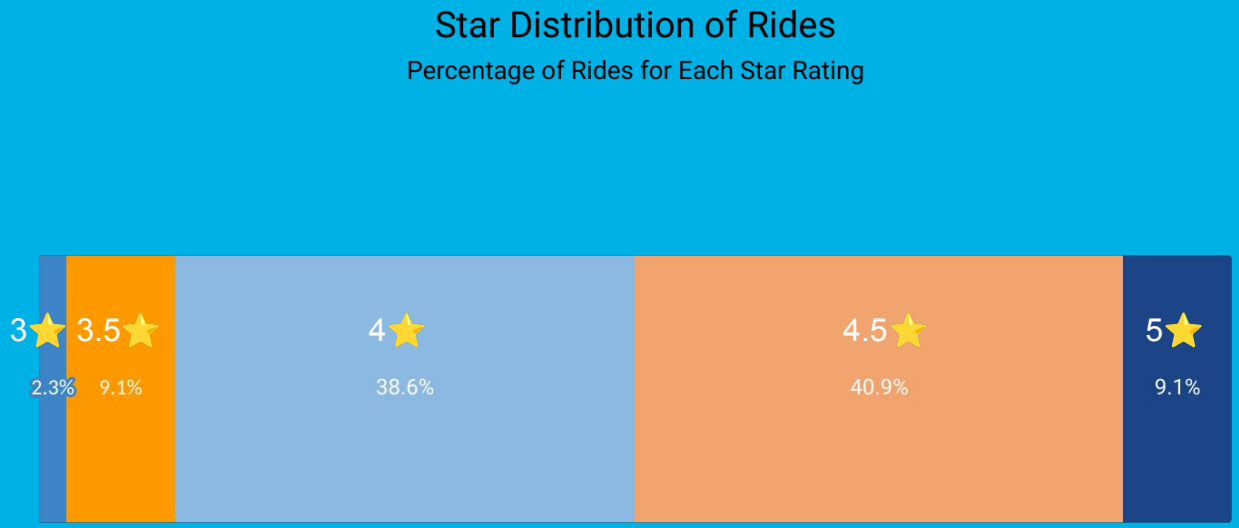
- My connection to the project
  - Annual Passholder at WDW
  - Avid Theme park attendee
- Dataset created by Lynne Passanisi on **Data.world**
- Data from 2018
  - New rides have since been opened
- Rides from all 4 Parks
  - Magic Kingdom
  - Hollywood Studios
  - Animal Kingdom
  - Epcot

Introduction
cont...

COLUMNS (27 in total)

- ★ Ride Name
- ★ Location (2 columns)
- ★ Ride Features (9 columns)
- ★ Fastpass Availability
- ★ “Classic” Ride Designation
- ★ Targeted Riders’ Age Group (5 columns)
- ★ Height Requirement
- ★ Ride Duration
- ★ Age of Ride (4 columns)
- ★ Ranking of Ride (via Travel & Leisure)
- ★ Star Rating (via Trip Advisor)

DATA VISUALIZATIONS

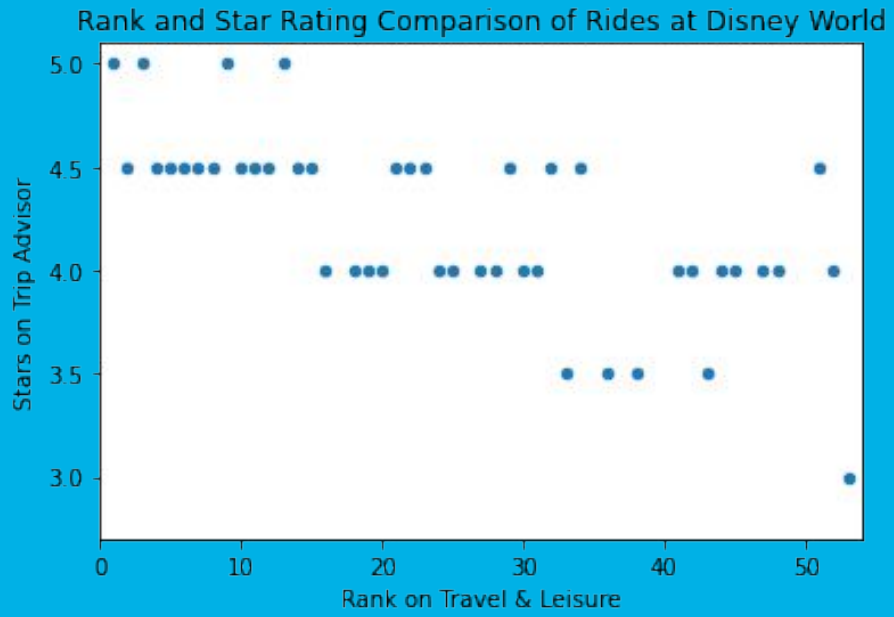


Distribution of star ratings:

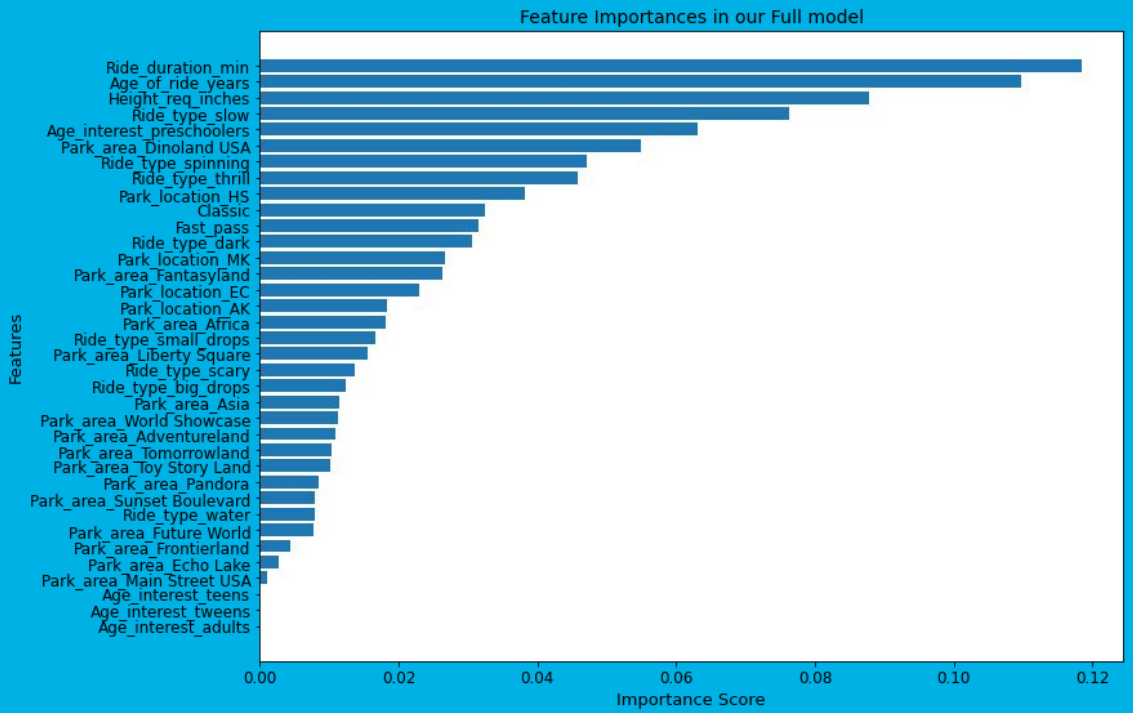
- All above 3 Stars
- 50% < or = 4 Stars
- 50% > 4 Stars

DATA VISUALIZATIONS

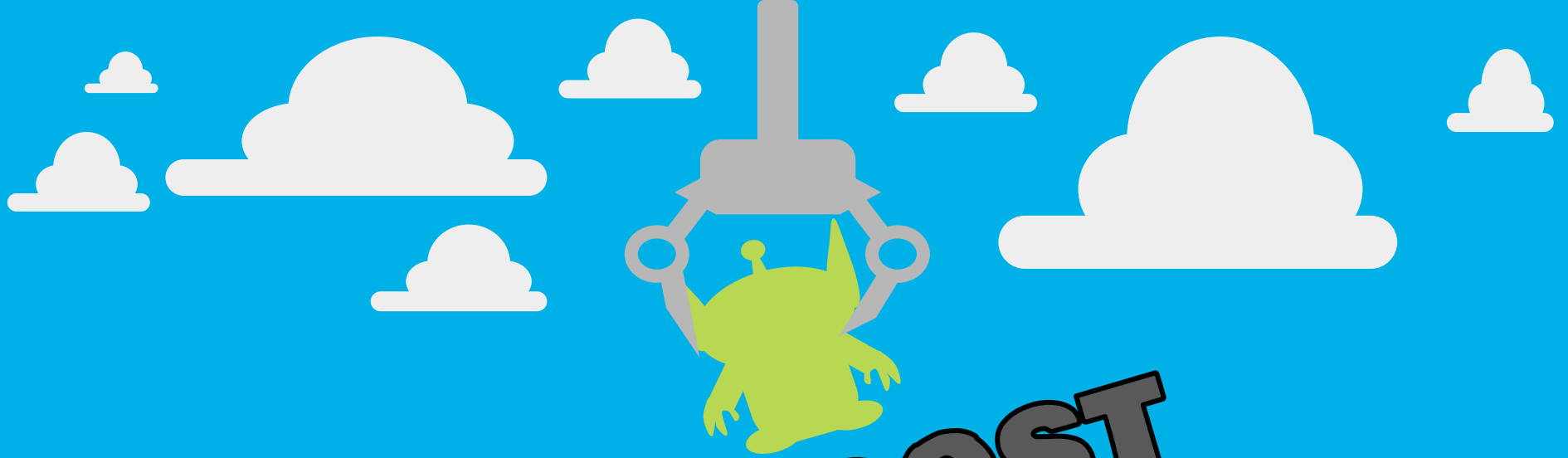
- ★ Rank is from 1 (best) - 53 (worst)
- ★ Star Ratings are on a scale of 1 (worst) - 5 (best)
  - All above 3★ as seen by last slide
- ★ Ranks and Stars does not 100% correlate: that is…
  - #1 is 5★ and #53 is 3★ BUT
  - the four 5★ rides are not ranked #1 - #4
  - The four 3.5★ rides are not ranked #48 - #52 etc.
- ★ Rank is an article/one opinion
- ★ Star Rating is an average of many reviews from attendees of the park



DATA VISUALIZATIONS



- ★ This chart shows the many features of our dataset
- ★ This is a ranking of how important each feature is to the model: i.e. how much does it actually affect the outcome?
- ★ High importance = more weight
- ★ Notice that some features had a score of 0
  - These were not included in the final model



XGBOOST

The Chosen Model

Model Evaluation

- ★ Baseline Accuracy **53%**
  - A baseline accuracy is if we guessed that everything was the majority option (lower starred rides).
- ★ The model predicted that **10/15** of the rides were low starred, when only **8/15** were
  - The model performs better for low-starred rides
    - It correctly identified about **88%** of what was low starred
  - Only slightly better than baseline in differentiating high starred rides
    - Identified **~57%** of the high starred rides correctly
  - Overall accuracy of **73%**

Recommendations

- ★ **Model is not yet ready for full implementation**
  - MUST expand the data set
  - more data (more rides) = can drastically improve this model!
    - Update to 2021/2022
    - Add rides from other popular amusement parks and/or other Disney Parks (Disneyland, Paris, Tokyo, Hong Kong, Shanghai)
- ★ So far we can determine what **features contribute for a successful/higher rated ride**
  - Avoid slower rides
  - Avoid spinning rides
  - Increase thrill rides

CREDITS

Presentation by...	**Mandy Sachs**
Email…	**mandy.sachs@gmail.com**
Professional Profile...	**Linkedin**
Portfolio…	**Github**

Slides Templates Designed by... **Matthew Meyer**