

POKÉMON! IS IT A LEGENDARY?

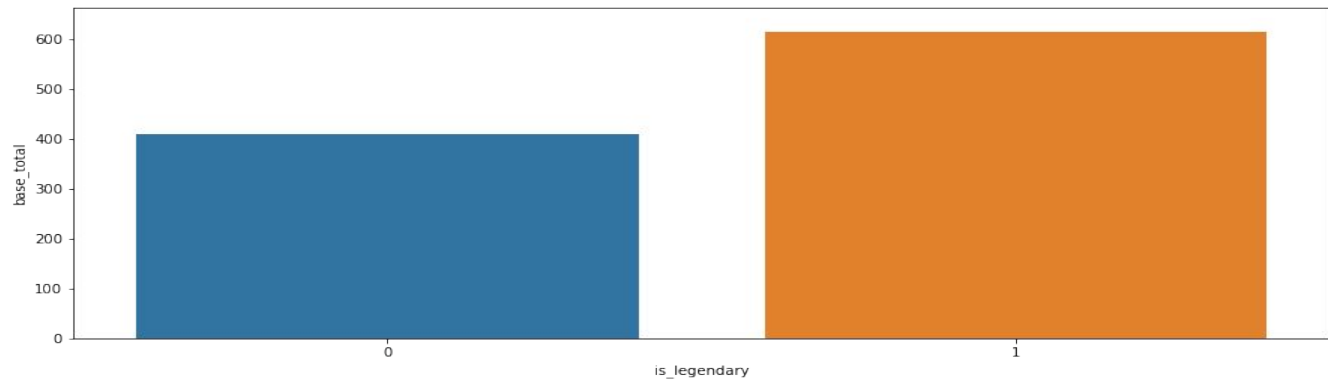
Bryce Liu, Kelly Zeng, Jerry Jiang (Mentor)

THE DATASET

- Average stats of every Pokemon until Generation 7 (Sun and Moon)
- 801 rows, 41 columns
 - abilities, against_water, base_egg_steps, attack, hp, speed, capture_rate, speed, name, generation, type1, type2, is_legendary, etc.
- Getting to know our data
 - `df.columns`
 - `df.dtypes`
 - `df.shape`
 - `df.describe`
 - `df.corr()`
 - Visualizations

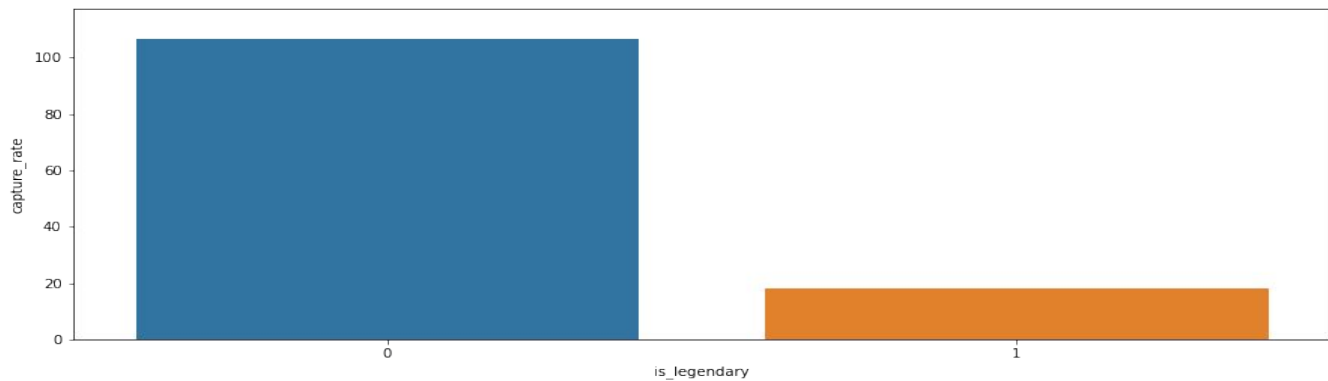
	against_bug	against_dark	against_dragon	against_electric	against_fairy	against_fight	against_fire	against_flying	against_ghost
against_bug	1.000000	0.230107	0.165430	-0.246943	0.239566	0.137902	0.202778	0.183343	0.129174
against_dark	0.230107	1.000000	0.140830	-0.015830	-0.301354	-0.357961	0.010527	-0.179697	0.672337
against_dragon	0.165430	0.140830	1.000000	-0.108928	0.439705	0.035237	-0.261570	0.064850	-0.049941
against_electric	-0.246943	-0.015830	-0.108928	1.000000	-0.089864	-0.102798	-0.279029	-0.111461	-0.073031
against_fairy	0.239566	-0.301354	0.439705	-0.089864	1.000000	0.157712	-0.169489	0.199862	-0.120806
against_fight	0.137902	-0.357961	0.035237	-0.102798	0.157712	1.000000	-0.076480	-0.318941	-0.546982
against_fire	0.202778	0.010527	-0.261570	-0.279029	-0.169489	-0.076480	1.000000	0.528127	0.050570
against_flying	0.183343	-0.179697	0.064850	-0.111461	0.199862	-0.318941	0.528127	1.000000	0.013204
against_ghost	0.129174	0.672337	-0.049941	-0.073031	-0.120806	-0.546982	0.050570	0.013204	1.000000
against_grass	0.079197	-0.006533	-0.037135	0.056209	0.052899	0.269157	-0.463634	-0.358467	0.029542
against_ground	-0.186841	-0.007660	-0.120042	-0.269444	-0.256504	0.358793	-0.260766	-0.465010	0.093727
against_ice	0.148176	-0.010763	0.350048	-0.328531	0.273650	-0.220239	0.186361	0.277935	-0.052531
against_normal	0.215689	-0.413632	0.142035	0.076699	0.149488	-0.006997	-0.055106	0.266780	-0.404780
against_poison	0.354255	-0.236919	-0.210199	-0.015769	0.146464	-0.189796	0.184334	0.484843	-0.105137
against_psychic	-0.463272	-0.230415	0.100153	-0.017592	-0.145238	-0.264938	-0.018519	0.375023	-0.084496
against_rock	-0.210522	0.011963	0.090184	0.417261	-0.205444	-0.240964	0.120747	0.161885	-0.061013
against_steel	0.055504	-0.119758	-0.227697	-0.187543	0.130323	0.165066	0.115992	0.001264	-0.002008
against_water	-0.254732	-0.001976	-0.096549	-0.297600	-0.218937	0.205249	-0.242635	-0.311997	0.009046
attack	-0.054175	-0.098849	0.138217	-0.104276	0.207526	0.149123	-0.005762	-0.002991	-0.034549
base_egg_steps	0.062133	0.187220	0.164773	-0.061970	0.120594	-0.006359	-0.044802	-0.091304	0.174115
base_happiness	0.009994	0.024155	-0.151915	0.030411	-0.209323	-0.088722	0.007441	0.056480	-0.037205
base_total	-0.012398	0.065446	0.069766	-0.017137	0.098948	0.048629	-0.061698	-0.106586	0.114179
defense	-0.036474	0.048039	-0.023794	-0.072433	0.001655	0.150424	0.037921	-0.127950	0.136490
experience_growth	0.035717	-0.008391	0.172547	-0.041584	0.146370	0.010407	-0.027136	0.019100	0.059262

VISUALIZATIONS

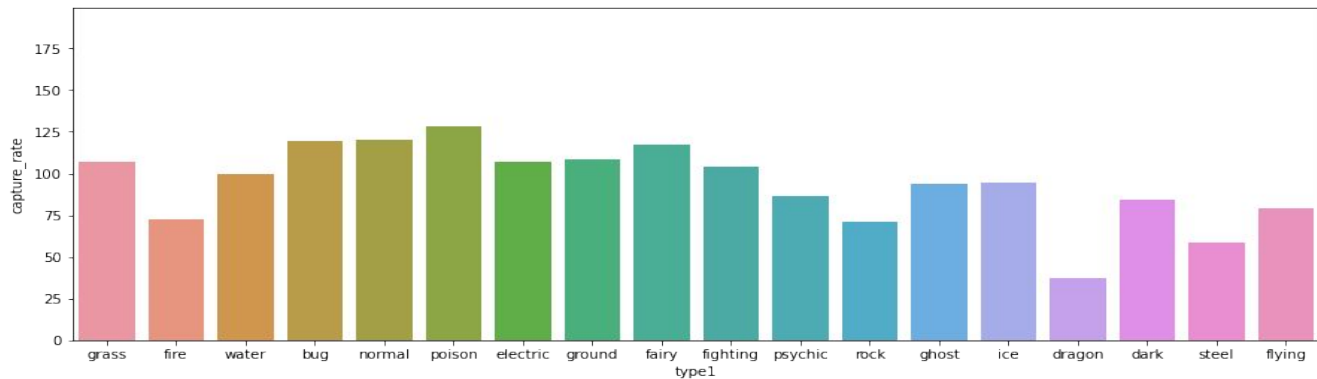
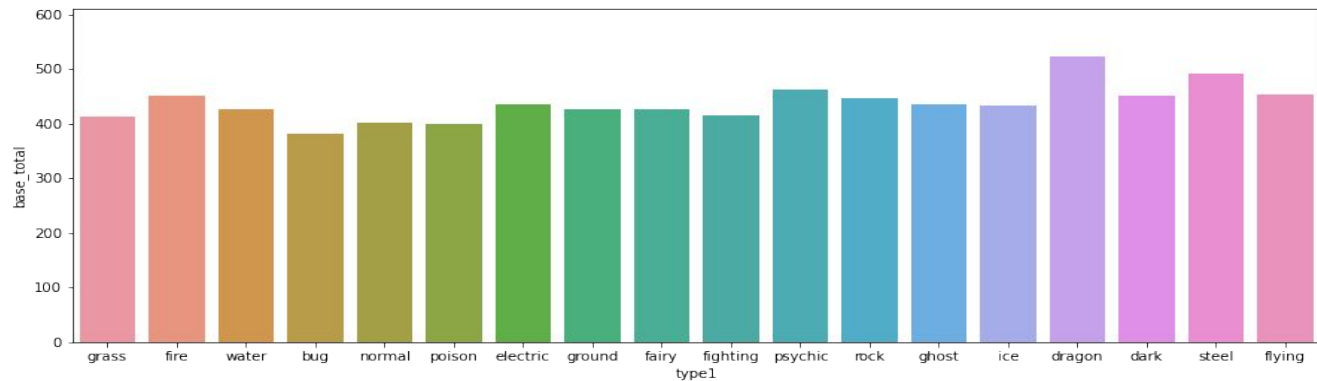


0 - non-legendary

1 - legendary



VISUALIZATIONS



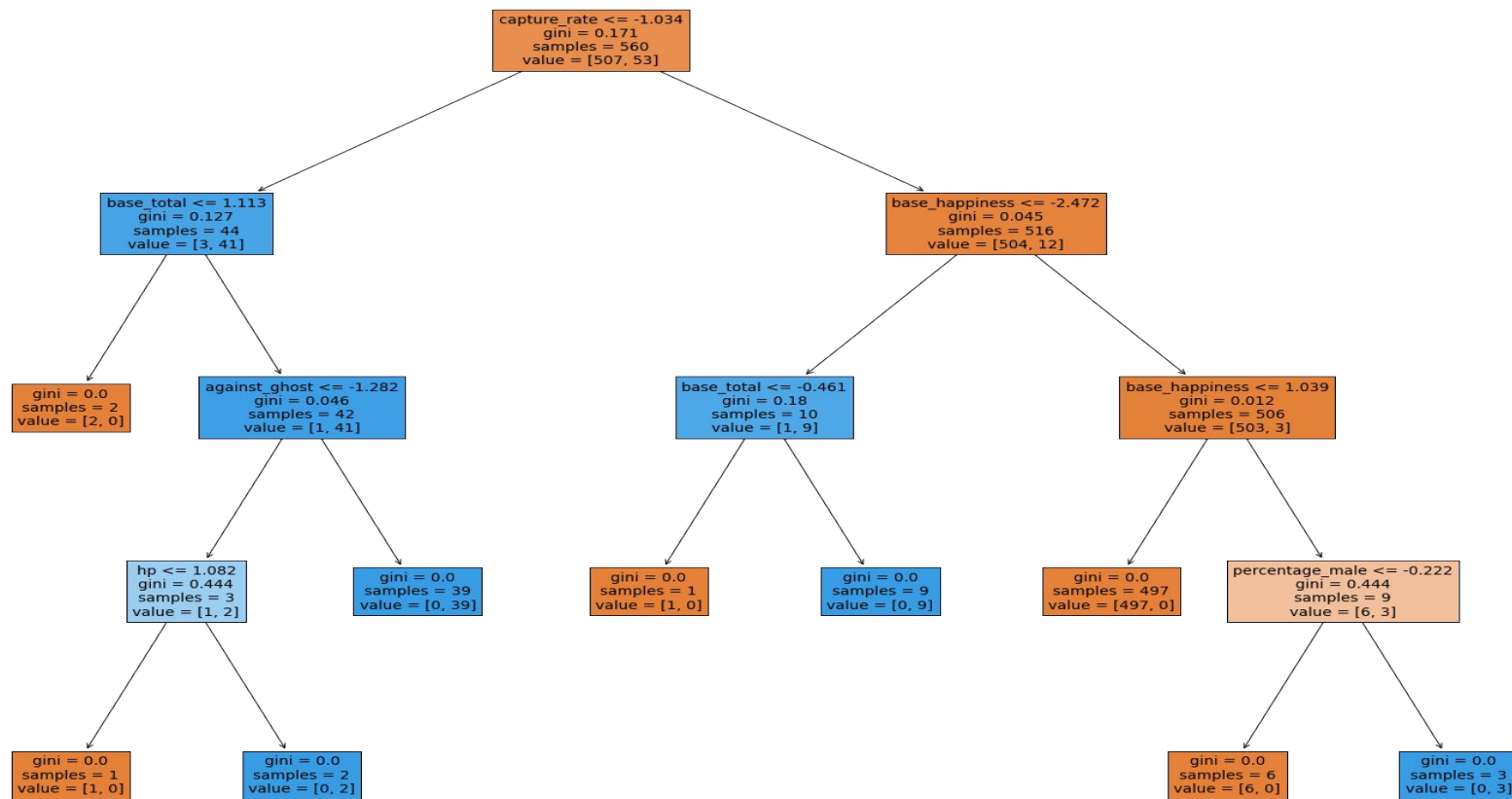
DATA CLEANING

- Replaced null values in `height_m`, `percentage_male`, `weight_kg` with average values grouped by `type1`
 - Mean imputation
- Dropped irrelevant columns
 - `abilities`, `japanese_name`, `generation`, `pokedex_number`, `name`, `classification`, `type1`, `type2`
- Transformed `capture_rate` from an object to a float
 - All columns are now numerical

CLASSIFICATION

- Determining if a Pokémon is legendary or not
 - Removed base_egg_steps since legendaries cannot hatch out of eggs
- Accuracy Score: 0.9917 (99.17%)
- Feature Importance (greatest to least):
 - capture_rate (0.697474)
 - base_happiness (0.183831)
 - base_total (0.056670)
 - percentage_male (0.041681)
 - hp (0.013894)
 - against_ghost (0.006451)

CLASSIFICATION



REGRESSION

- Predicting the capture rate of a Pokémon
- Linear Regression
 - MSE: 3299.7881454768644
- LASSO
 - MSE: 3102.9517871682874
 - Kept base_total, hp, percentage_male
- Ridge
 - MSE: 3273.0234998175033
- Elastic Net
 - MSE: 3103.3369296640058
- KNN (k = 10)
 - MSE: 5264.535269709543
 - Prediction Accuracy: 0.3941908713692946

CONCLUSIONS

- Our classification model is highly accurate
 - capture_rate is the most important variable (besides base_egg_steps) in determining if a Pokémon is legendary or not, followed by base_total
- LASSO Regression performs the best
 - Smallest MSE
 - base_total, hp, and percentage_male are strongly associated with the capture rate of a Pokémon