Predicting Game's Sales

By Ghanim Alghanim, Ali Altamimi

Gathering Data

Web Scraping

- The dataset from collected VGChartz
- ~61,000 observation and 18 features.
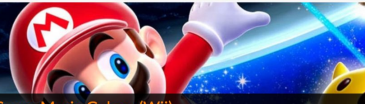- Some of the features are platform, genre, score, user vote count… etc.

Data Cleaning

Data Cleaning & Feature Engineering
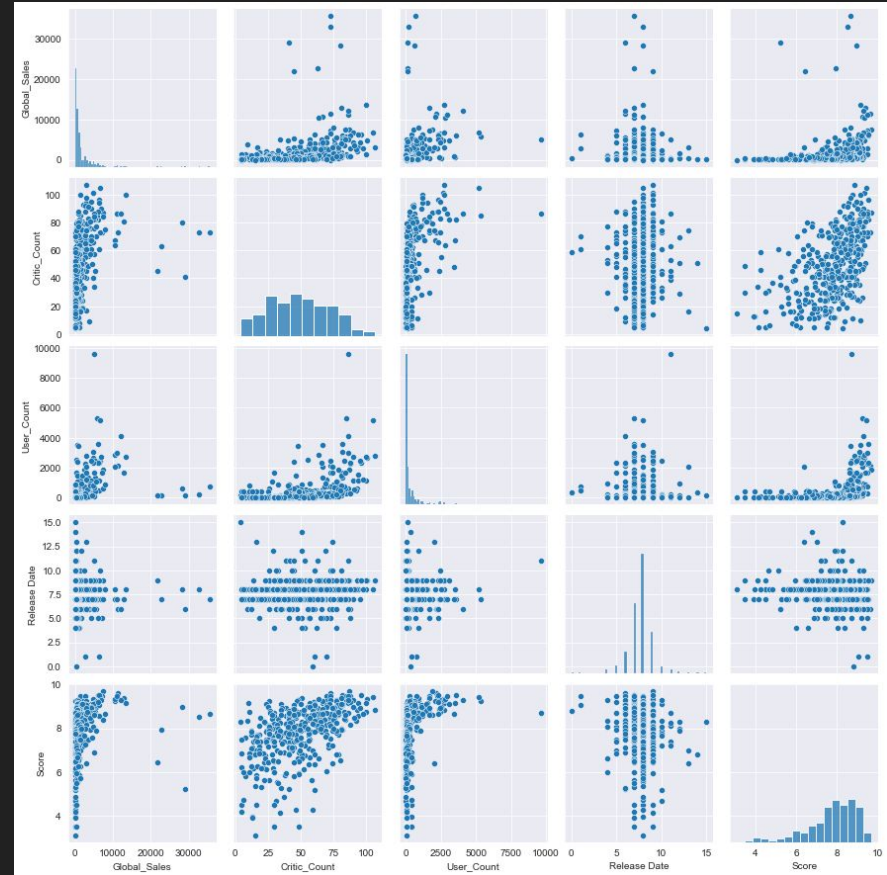
Data Cleaning & Feature Engineering

- ~400 observation.
- ~38 features.
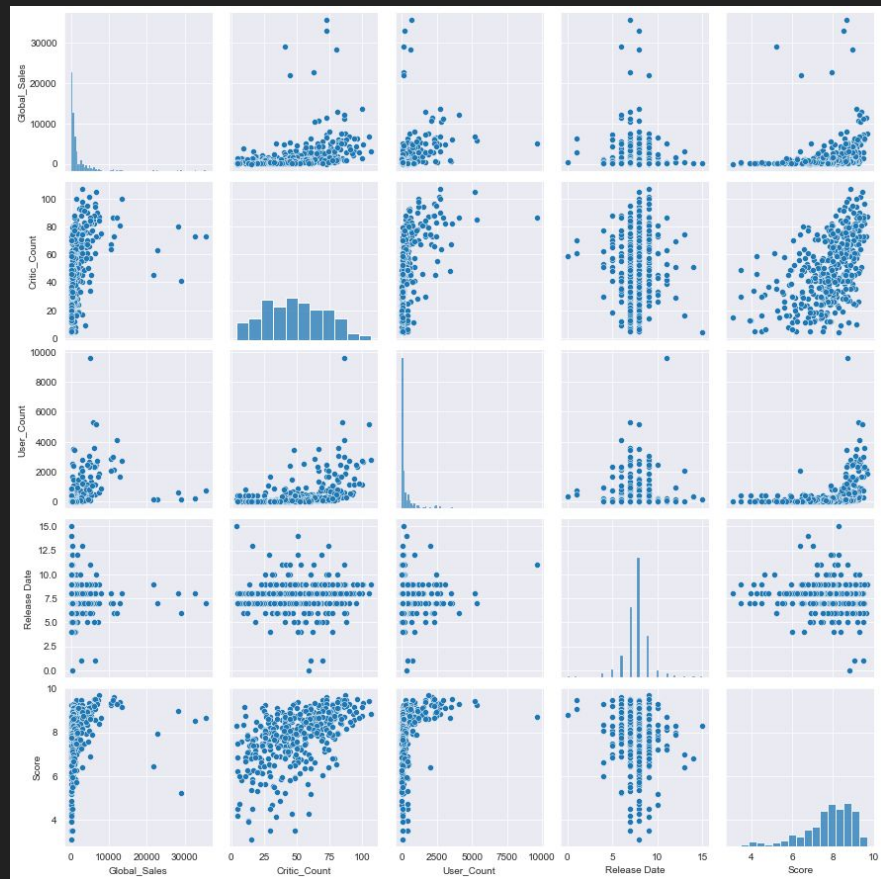
Creating a Model

Splitting The Data Frame

- No Categorical
- Genre
- Console
- Rating
- Console & Rating
- Genre & Console
- Genre & Rating
- All Categorical

Models used

- Linear Regression
- Lasso Regression
- Ridge Regression

	Linear Score	Lasso Score	Ridge Score
No Categorical	0.172062	0.172069	0.172088
Genre	0.160681	0.160806	0.163956
Console	0.242307	0.242381	0.244139
Rating	0.240552	0.240547	0.240355
Console & Rating	0.302434	0.302504	0.304803
Genre & Console	0.221529	0.222290	0.233919
Genre & Rating	0.193233	0.193353	0.198735
All Categorical	0.236200	0.236593	0.253576

Prediction

```
big_new_df_Genre_Rating.columns
```

```
Index(['Global_Sales', 'Critic_Count', 'User_Count', 'Release Date', 'Score',
       'Genre_Action', 'Genre_Action-Adventure', 'Genre_Adventure',
       'Genre_Fighting', 'Genre_Misc', 'Genre_Music', 'Genre_Party',
       'Genre_Platform', 'Genre_Puzzle', 'Genre_Racing', 'Genre_Role-Playing',
       'Genre_Shooter', 'Genre_Simulation', 'Genre_Sports', 'Genre_Strategy',
       'Rating_E', 'Rating_E10+', 'Rating_M', 'Rating_T'],
      dtype='object')
```

[0,0,22,9.9,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1]

Prediction Result = 90m

Thank you

Questions ?