Video Game Analysis Through Sales, Reviews, and Regional Preferences

Sharon Yin, Kyra Zhu
4/30/22

Introduction and Questions

- Video game sales
- Questions
  - What are the relationship/correlation between critic scores and total sales?
  - How does the distribution of sales vary across different regions?
  - Is there any correlation between game sales and pricing?

Datasets

- Datasets
  - Video game sales data
    - https://www.kaggle.com/datasets/asaniczka/video-game-sales-2024
  - Steam games dataset
    - https://www.kaggle.com/datasets/bernsn/steam-games-2023-cleaned
- Clean data
  - Drop na data, turn all the reviews in steam to numerical numbers, drop $ signs, drop duplicates
- Merged dataset
  - Focus on PC games

```
[25] games = pd.read_csv("./vgchartz-2024.csv")
     games = games[["title", "console", "genre", "critic_score", "total_sales", "na_sales", "jp_sales", "pal_sales", "other_sales"]]
     games
```

	title	console	genre	critic_score	total_sales	na_sales	jp_sales	pal_sales	other_sales
0	Grand Theft Auto V	PS3	Action	9.4	20.32	6.37	0.99	9.85	3.12
1	Grand Theft Auto V	PS4	Action	9.7	19.39	6.06	0.60	9.71	3.02
2	Grand Theft Auto: Vice City	PS2	Action	9.6	16.15	8.41	0.47	5.49	1.78
3	Grand Theft Auto V	X360	Action	NaN	15.86	9.06	0.06	5.33	1.42
4	Call of Duty: Black Ops 3	PS4	Shooter	8.1	15.09	6.18	0.41	6.05	2.44

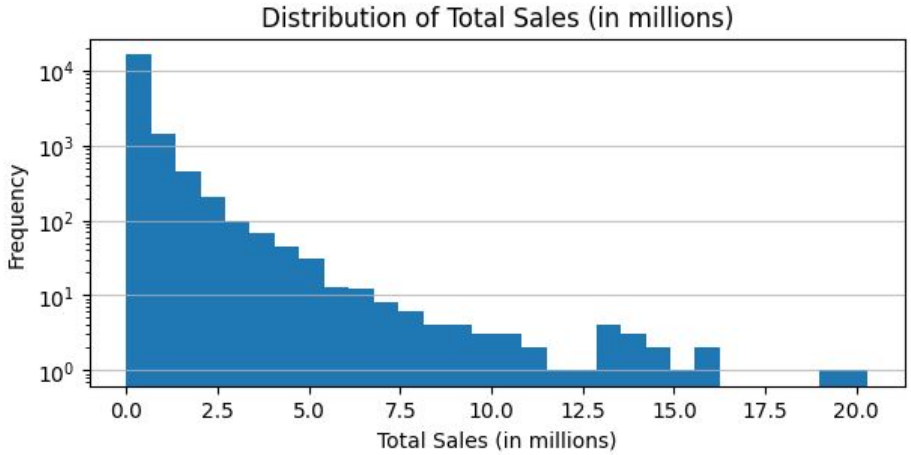
	title	price	allReviews
0	Ori and the Will of the Wisps	9.89	9.0
1	Flashing Lights - Police, Firefighting, Emerge...	8.49	8.0
2	Thronefall	5.24	9.0
3	DRAGON QUEST® XI S: Echoes of an Elusive Age™ ...	23.99	8.0
4	UNDYING	13.99	6.0

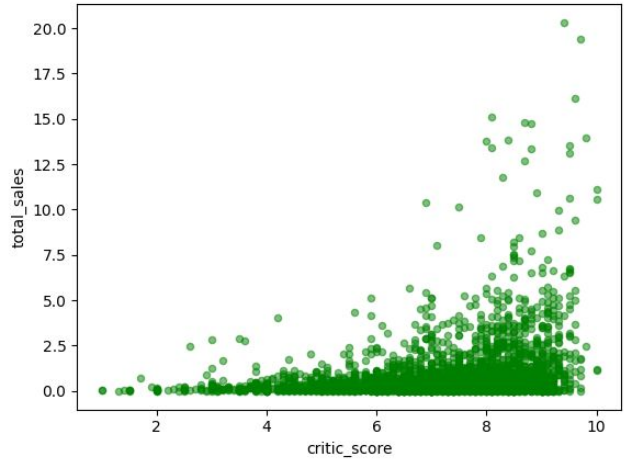
```
pc_merged_data
```

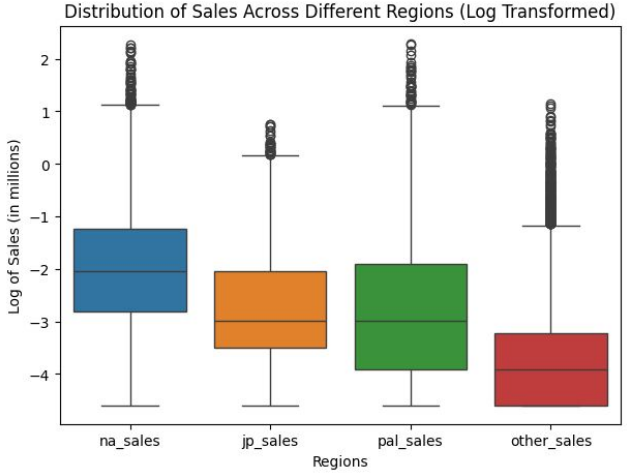
	title	price	allReviews	console	genre
0	Ori and the Will of the Wisps	9.89	9.0	PC	Action-Adventure
2	Thymesia	14.99	8.0	PC	Role-Playing
4	Ready or Not	37.49	6.0	PC	Shooter
6	Destiny 2	14.99	6.0	PC	Shooter
7	Hogwarts Legacy	35.99	8.0	PC	Role-Playing

Graphs and Statistics

Exploratory analysis to show important features







Hypothesis

- Objective: predict game sales based on different features
- Features: console, genre, critic score, release date
- Target: total sales
- Processing: pipeline integrates preprocessing and model training
- Model/methods: decision tree RandomForestRegressor model, KNN model, cross validation
- Data splitting: 80% training, 20% testing

Conclusion

Result:

- Decision tree RandomForestRegressor:
  - MSE: 1.623
- KNN:
  - MSE: 1.553
- Cross validation
  - Mean RMSE for 5 folds: 1.302

Conclusion:
Cross-Validation with Random Forest achieves the lowest Mean RMSE, indicating superior predictive accuracy.
KNN performs slightly better than Decision Tree RandomForestRegressor based on MSE.
For predicting video game sales, Cross-Validation with Random Forest is recommended for its better overall performance.