

# Predicting Video Game Sales in North America: Data and EDA

Date of Presentation: 11-05-2024

Student Name: Nirja Shah

Student's Pace Email Address: ns53074n@pace.edu

Class Name: Practical Data Science

Program Name: MS in Data Science

Seidenberg School of Computer Science and Information Systems

Pace University

# Agenda

---

- Executive summary
- Project plan recap
- Data
- Exploratory data analysis
- Modeling methods
- Findings
- Recommendations and technical next steps

# Executive Summary

---

- Problem:

How do we optimize game promotion strategy in the North American region using the historical sales data available?

- Solution:

The goal is to identify a prediction model to estimate sales in the North American region based on publisher, platform and genre using previous data on game sales. With this knowledge, the business can focus its marketing efforts on the platforms and genres that are most likely to succeed.

# Project Plan

---

Deliverable	Due Date	Status
Data and EDA	11-05-2024	Completed
Methods, Findings and Recommendations	11-12-2024	Not Started
Final Presentation	12-03-2024	Not Started

Data

# Data

---

- Data Source: Video Game Sales 2024 from [Kaggle](#)
- Sample Size: 64016 Rows
- Attributes Used:
  - Console
  - Release Year
  - Publisher
  - Developer
  - Title
  - Sales by Region (North America)
- Exclusions:
  - Critic Score are excluded as they do not serve any purpose in the modeling

# Data (continued)

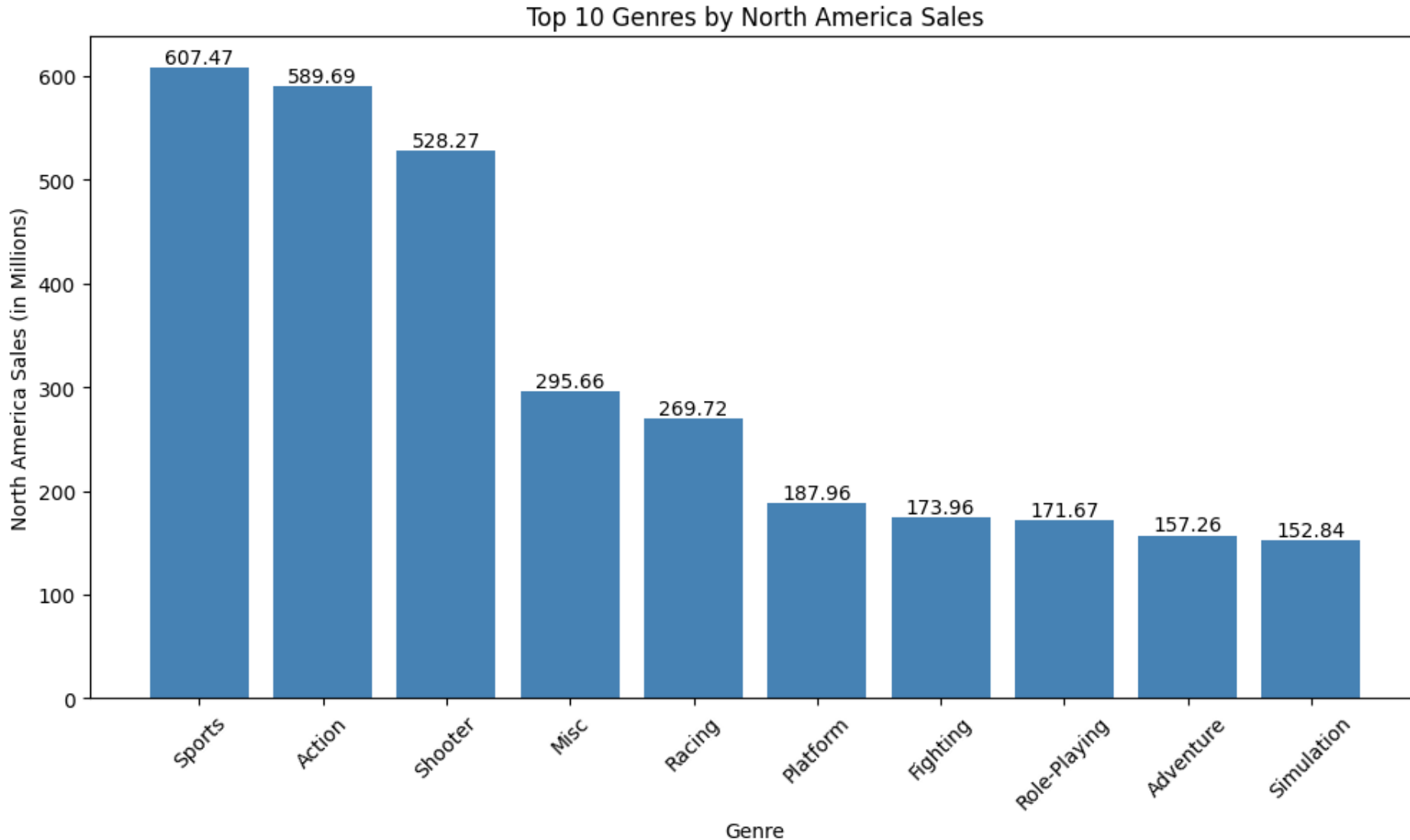
---

- Time Period: The data covers video game sales from as back as 1970 to as recent as January 2024.
- Assumption:
  - Since the majority of game shops and platforms submit sales data for North America, we assume that the dataset represents a representative sample of video game sales in this region. This means that our study accurately reflects North American sales patterns.

# Exploratory Data Analysis



# Top 10 Genres

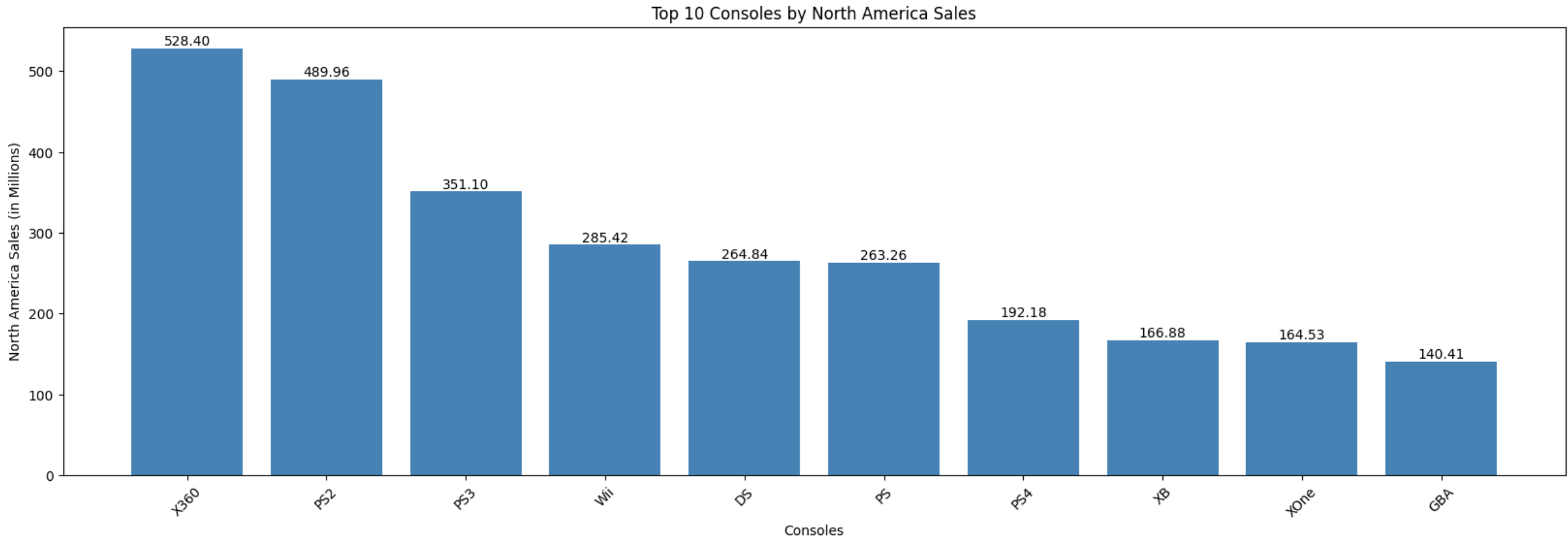


## Key Points:

Sports, Action, and Shooter genres dominate North American game sales, while Racing, Platform, and other genres contribute moderately.

Note: For sales data across all genres, please refer to [slide 16](#) appendix

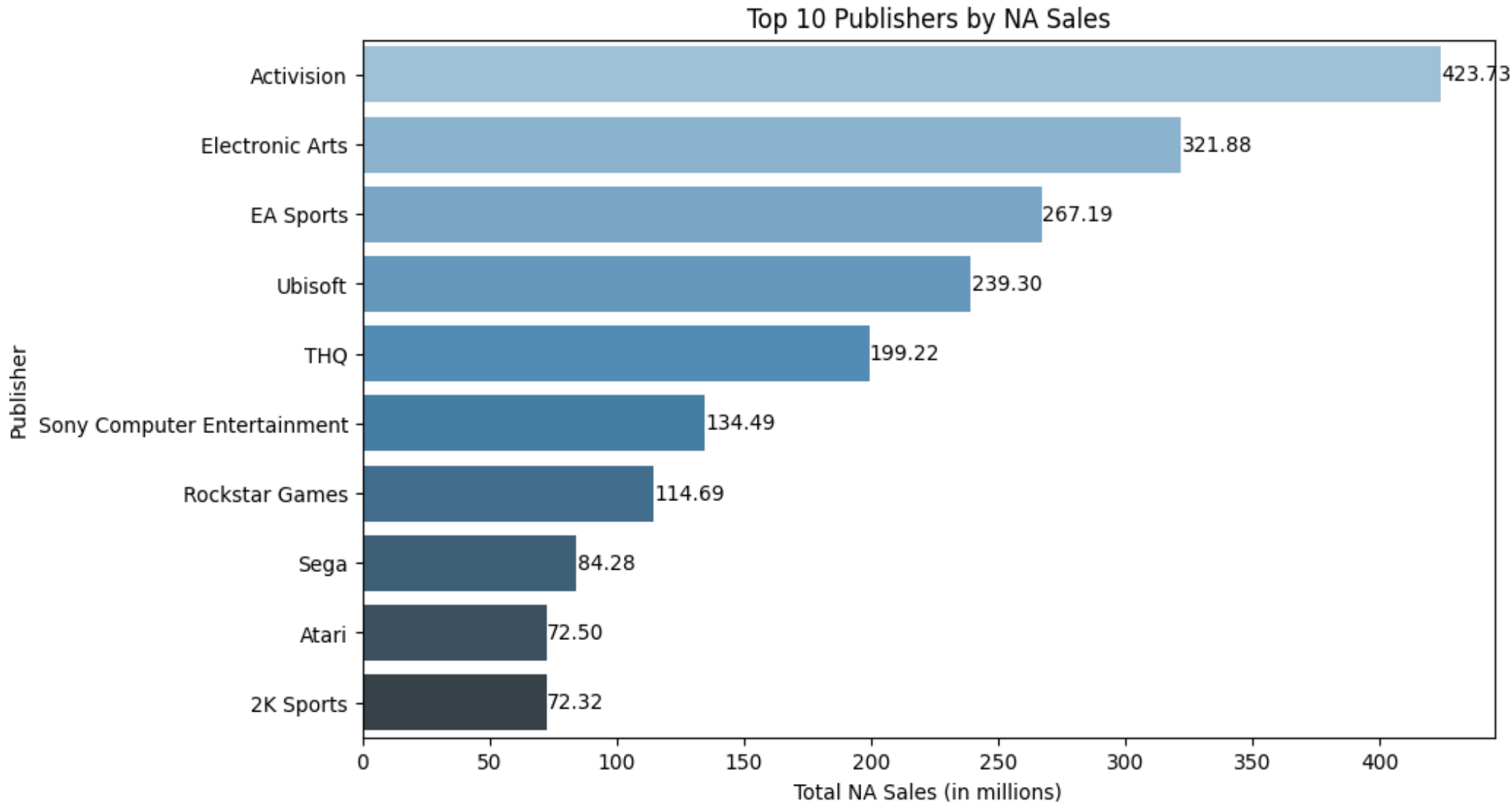
# Top 10 Consoles



Key Points: The Xbox 360 and PS2 lead North American console sales, with other popular consoles like the PS3, Wii, and DS following behind.

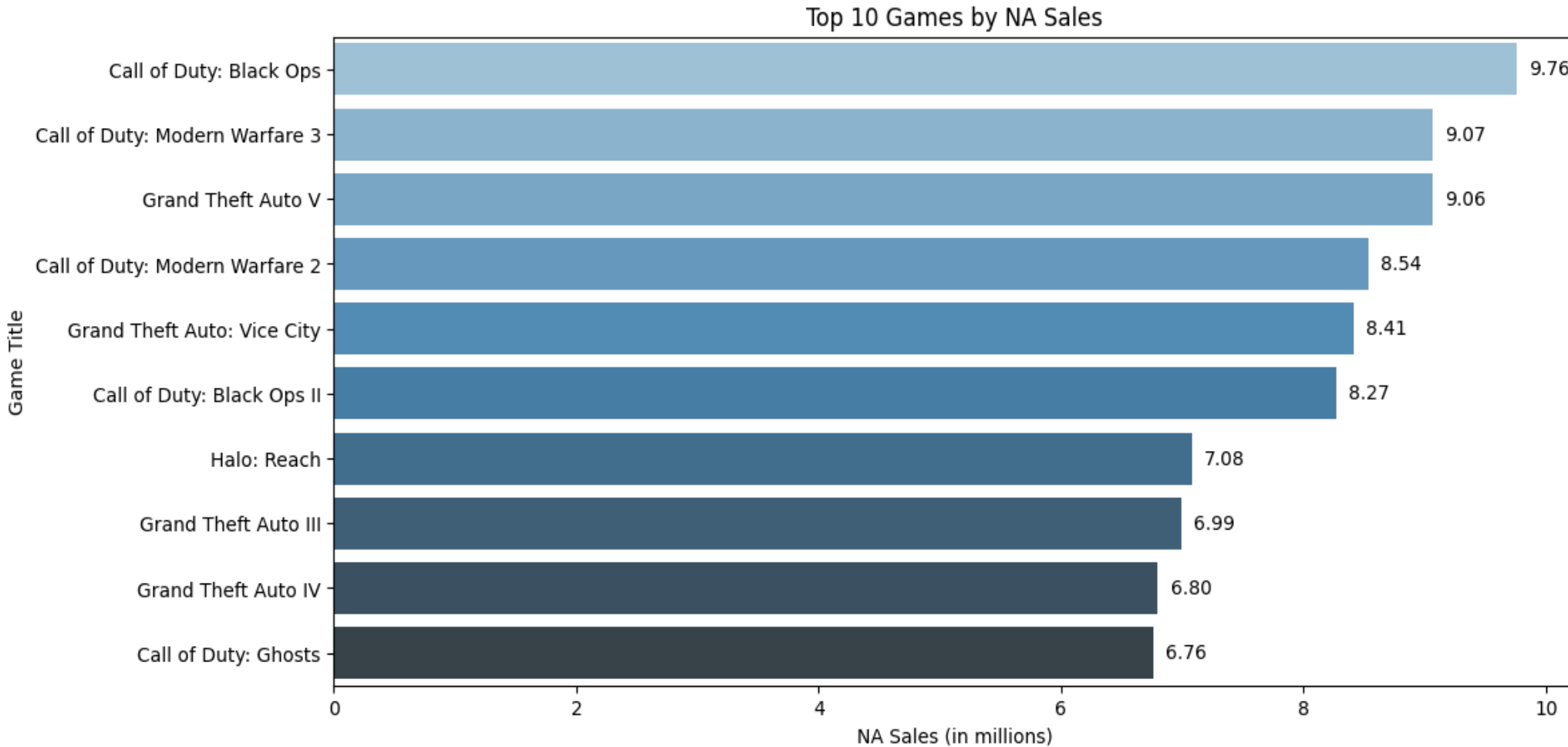
Note: For sales data across all consoles, please refer to [slide 17](#) appendix

# Top 10 Publisher



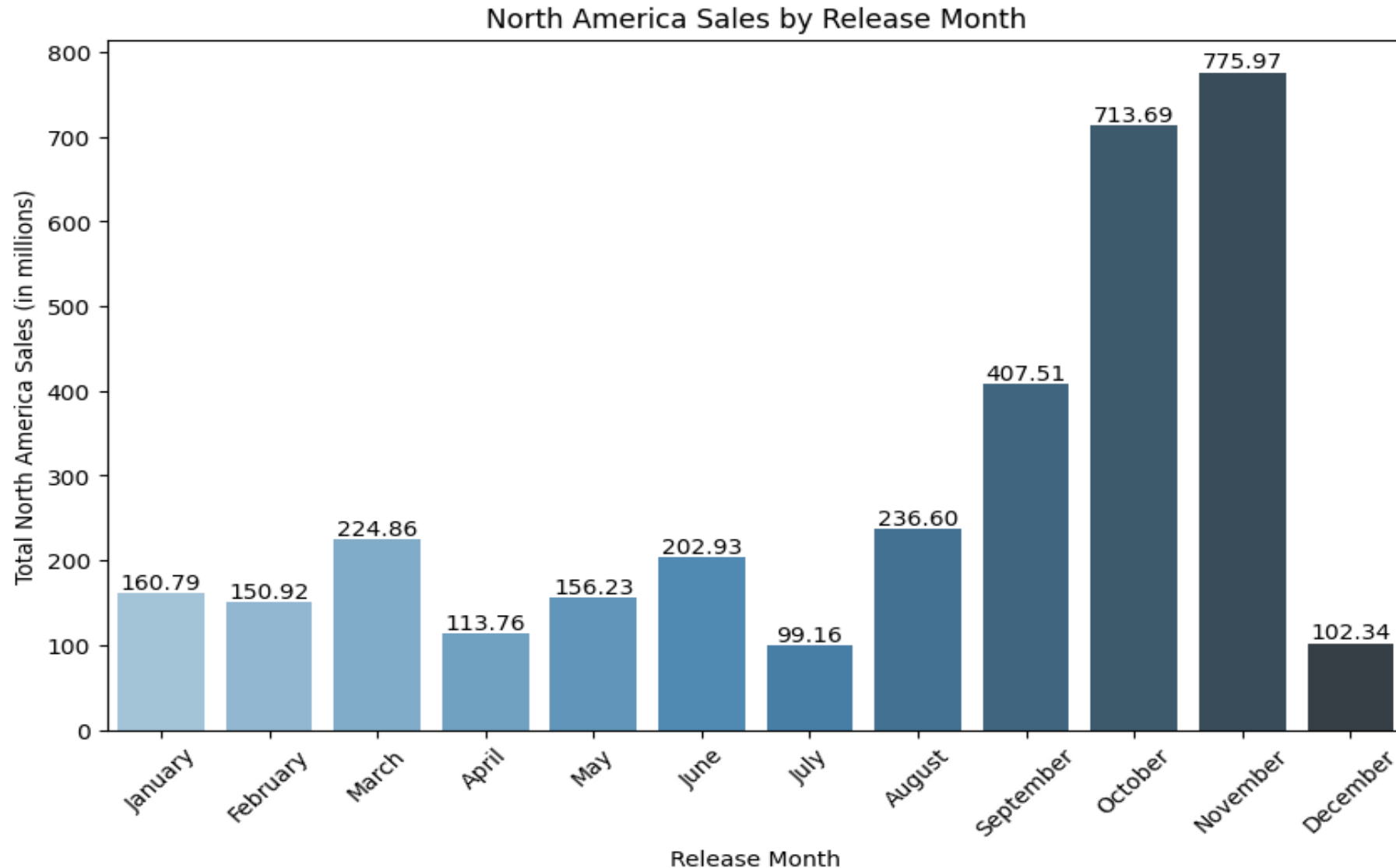
**Key points:**  
The chart shows that Activision and Electronic Arts (EA) dominate North American video game sales, with Activision at the top (423.73 million) and EA, including EA Sports, following closely.

# Top 10 Games



Key Points:  
The *Call of Duty* and *Grand Theft Auto* franchises dominate the top 10 NA sales chart, with *Call of Duty: Black Ops* leading at 9.76 million units.

# Sales by Month



## Key Point:

This chart shows strong seasonality in NA game sales, peaking in October (713.69M) and November (775.97M) due to holiday shopping.

Sales dip in the summer, with July being the lowest (99.16M).

March and September also see moderate increases, likely due to release cycles.

# Appendix

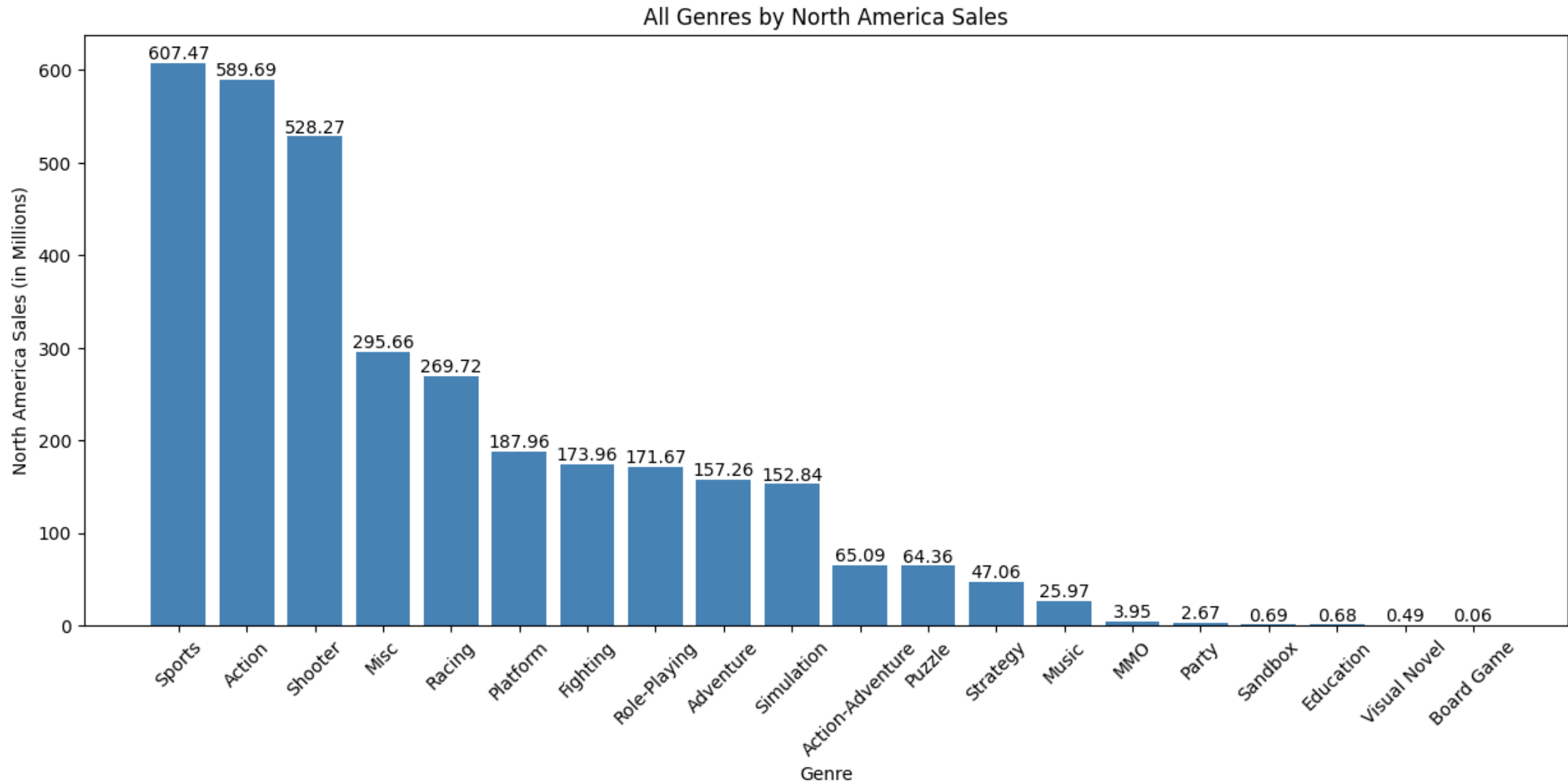
# Data Features

---

- All features 14 features in the dataset are as follows:

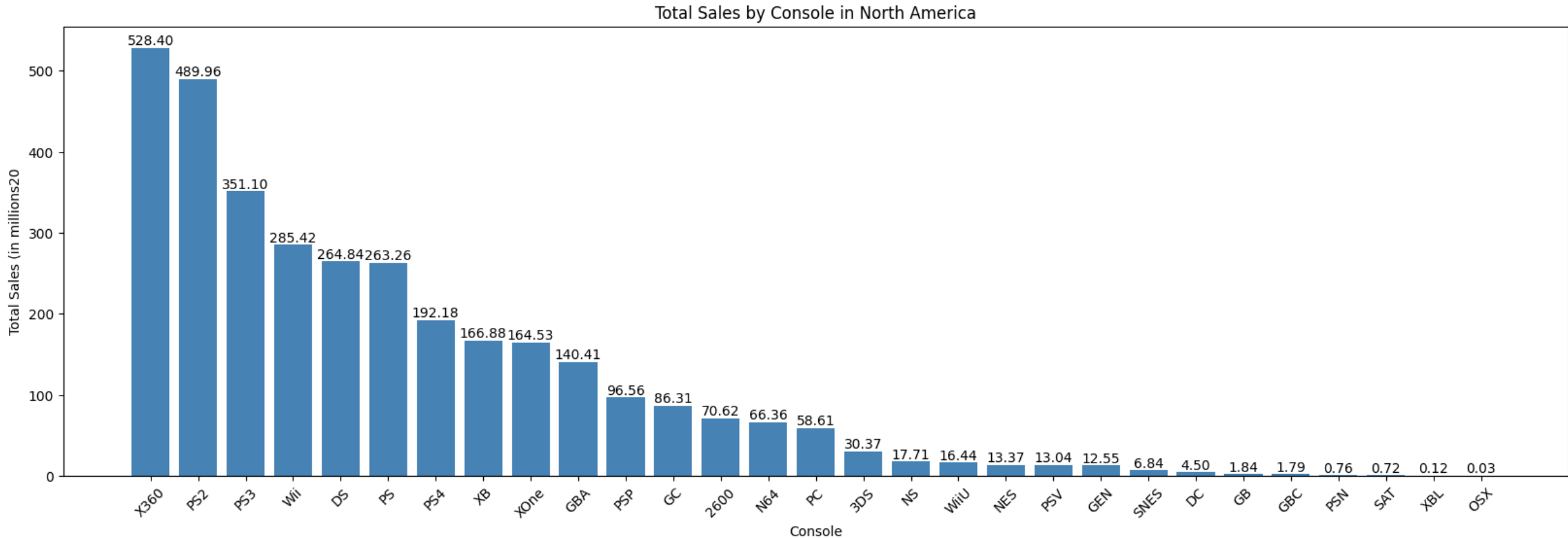
1. **Image:** Cover image of the game.
2. **Title:** Name of the game.
3. **Publisher:** Name of the publisher.
4. **Developer:** Name of the Developer.
5. **Console:** Console on which this game was released on.
6. **Genre:** Genre of the game.
7. **Critic score:** Metacritic score (out of 10).
8. **Total Sales:** Global sales of copies in millions.
9. **North America sales:** North American sales of copies in millions.
10. **Japan Sales:** Japan sales of copies in millions.
11. **PAL Sales:** Europe/Africa sales of copies in millions.
12. **Other Sales:** Other sales of copies in millions.
13. **Release Date:** Date of the game released.
14. **Last Updated:** Date the data was last updated.

# All Genres





# All Console



Note: There are total 81 unique consoles, here about 29 are only visible because all the other genres had less than 0.1 sales.