



Video games sales regression

Poojitha Annabathula
Assyl Salah

Zuzanna Kowalczyk
Viktor Sas



TABLE OF CONTENTS



Introduction

Why We Chose This Topic

Problem Overview

Our Approach

Methodology

Tools and Libraries

Team Responsibilities

Work split





01

Introduction



└ Why We Chose This Topic ─



Real-World Impact

Predicting game sales can support strategic decisions in marketing and development



Open & Reproducible Dataset

Data from Kaggle includes clear structure and all preprocessing steps, supporting reproducibility



Popular & Growing Industry

The video game industry is one of the largest and fastest-growing entertainment sectors



Rich Features for Analysis

Variables like critic scores, user reviews, genre, and platform allow deep exploration





02

Problem Overview





Our Approach



Environment Setup

Requirements text file
Set Random Seed
Virtual Environment
Include README.md

Handling Missing data

Dropping all NAs

Feature Engineering

Filtering data
Removing outliers
Creation of new variables

Data Visualization

Visualize trends and distributions (e.g., global sales histograms) to explore data patterns

Regression Models

Linear Regression
Ridge
Lasso

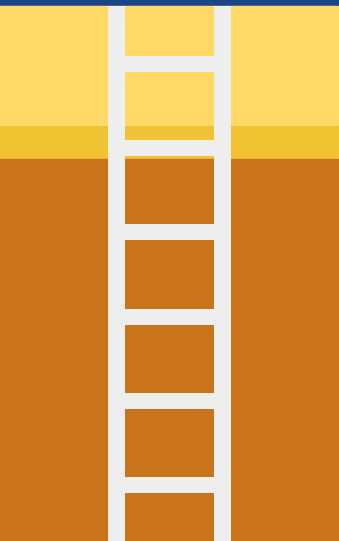
Model Evaluation

RMSE
 R^2 Score



03

Methodology





Tools and Libraries



Tool / Libraries	Description
Pandas & NumPy	For data loading, cleaning and manipulation
Matplotlib & Seaborn	Used for creating histograms, scatter plots, and other insightful visualization
Scikit-learn	Tool for model building, data splitting, and performance evaluation
Jupyter Notebook	Interactive environment for step-by-step development and reproducibility
GitHub	For version control, collaboration, and sharing a reproducible research pipeline



04

Team Responsibilities





Responsibilities



Assyl

Preparing
Presentation,
Environment Setup,
documentation and
Reproducibility



Zuzanna

Data Visualization



Data preparation
and cleaning

Viktor



Data Modeling

Poojitha

