Games Sales Prediction Using Time Series Analysis

Presented by

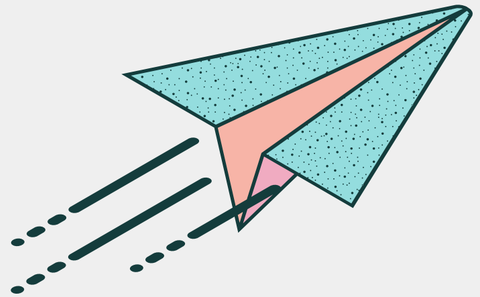
Aidar Ospankulov

Olzhas Assetuly

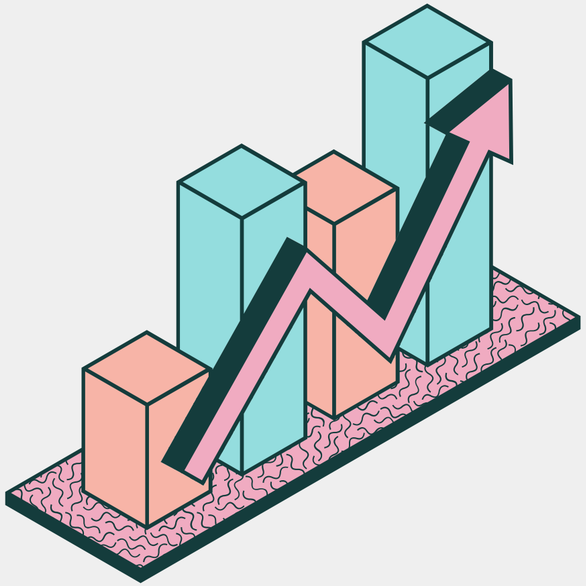
Timur Seitov

Eldar Batkenov

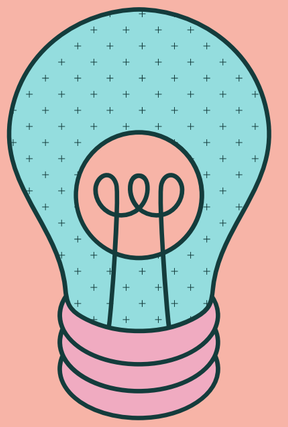
Purpose



Predict future video game sales using historical data to aid in strategic planning and marketing.

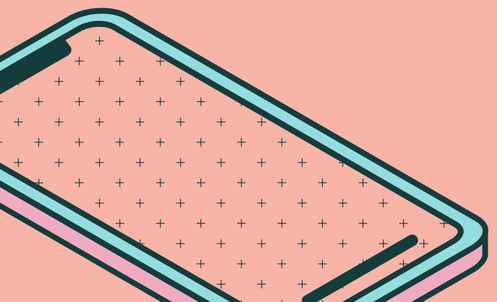


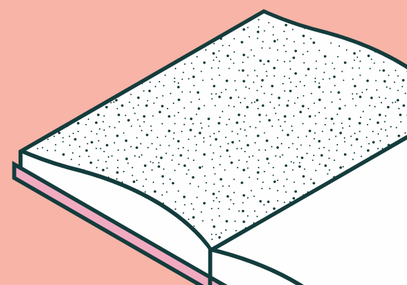
Dataset preparation



The dataset preparation includes:

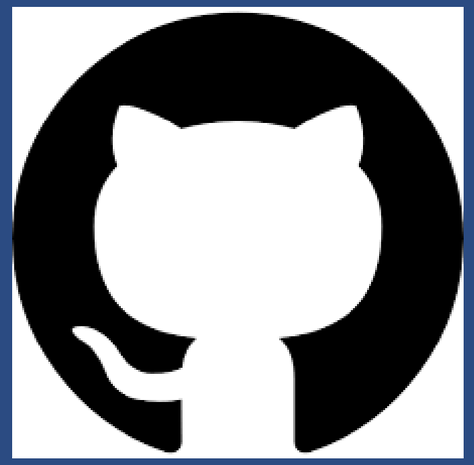
1. Handle Missing Values
2. Duplicates Removal
3. Split the dataset into train and test set and use them for model preparation





Data Collection

- VGChartz
- Kaggle



Main Features

- Game release date
- Sales
- Genre and platform
- Additional factors: reviews, user and critic ratings, sales region

Dataset

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	img_	title	console	genre	publisher	developer	critic_scor	total_sales	na_sales	jp_sales	pal_sales	other_sale	release_date	last_update
2	/game	Grand Theft Auto V	PS3	Action	Rockstar G	Rockstar N	9.4	20.32	6.37	0.99	9.85	3.12	17.09.2013	
3	/game	Grand Theft Auto V	PS4	Action	Rockstar G	Rockstar N	9.7	19.39	6.06	0.6	9.71	3.02	18.11.2014	03.01.2018
4	/game	Grand Theft Auto: Vice City	PS2	Action	Rockstar G	Rockstar N	9.6	16.15	8.41	0.47	5.49	1.78	28.10.2002	
5	/game	Grand Theft Auto V	X360	Action	Rockstar G	Rockstar		15.86	9.06	0.06	5.33	1.42	17.09.2013	
6	/game	Call of Duty: Black Ops 3	PS4	Shooter	Activision	Treyarch	8.1	15.09	6.18	0.41	6.05	2.44	06.11.2015	14.01.2018
7	/game	Call of Duty: Modern Warfare	X360	Shooter	Activision	Infinity Wa	8.7	14.82	9.07	0.13	4.29	1.33	08.11.2011	
8	/game	Call of Duty: Black Ops	X360	Shooter	Activision	Treyarch	8.8	14.74	9.76	0.11	3.73	1.14	09.11.2010	
9	/game	Red Dead Redemption 2	PS4	Action-Adv	Rockstar G	Rockstar G	9.8	13.94	5.26	0.21	6.21	2.26	26.10.2018	02.11.2018
10	/game	Call of Duty: Black Ops II	X360	Shooter	Activision	Treyarch	8.4	13.86	8.27	0.07	4.32	1.2	13.11.2012	07.04.2018
11	/game	Call of Duty: Black Ops II	PS3	Shooter	Activision	Treyarch	8	13.8	4.99	0.65	5.88	2.28	13.11.2012	07.04.2018
12	/game	Call of Duty: Modern Warfare	X360	Shooter	Activision	Infinity Wa	9.5	13.53	8.54	0.08	3.63	1.28	10.11.2009	
13	/game	Call of Duty: WWII	PS4	Shooter	Activision	Sledgehan	8.1	13.4	4.67	0.4	6.21	2.12	03.11.2017	31.12.2017
14	/game	Call of Duty: Modern Warfare	PS3	Shooter	Activision	Infinity Wa	8.8	13.35	5.54	0.49	5.78	1.54	08.11.2011	
15	/game	Grand Theft Auto III	PS2	Action	Rockstar G	DMA Desig	9.5	13.1	6.99	0.3	4.51	1.3	23.10.2001	
16	/game	Call of Duty: Black Ops	PS3	Shooter	Activision	Treyarch	8.7	12.67	6.01	0.48	4.4	1.78	09.11.2010	
17	/game	FIFA 18	PS4	Sports	EA Sports	EA Vancou	8.3	11.8	1.27	0.15	8.64	1.73	29.09.2017	02.04.2018
18	/game	Grand Theft Auto IV	X360	Action	Rockstar G	Rockstar N	10	11.09	6.8	0.14	3.11	1.04	29.04.2008	
19	/game	FIFA 17	PS4	Sports	Electronic	EA Canada	8.9	10.94	1.26	0.12	7.95	1.61	27.09.2016	28.03.2018
20	/game	Call of Duty: Modern Warfare	PS3	Shooter	Activision	Infinity Wa	9.5	10.61	4.99	0.38	3.66	1.59	10.11.2009	

Time series models

SARIMA (Seasonal Auto-Regressive Integrated Moving Average)

- Combines different techniques to work with data that doesn't have a clear pattern.

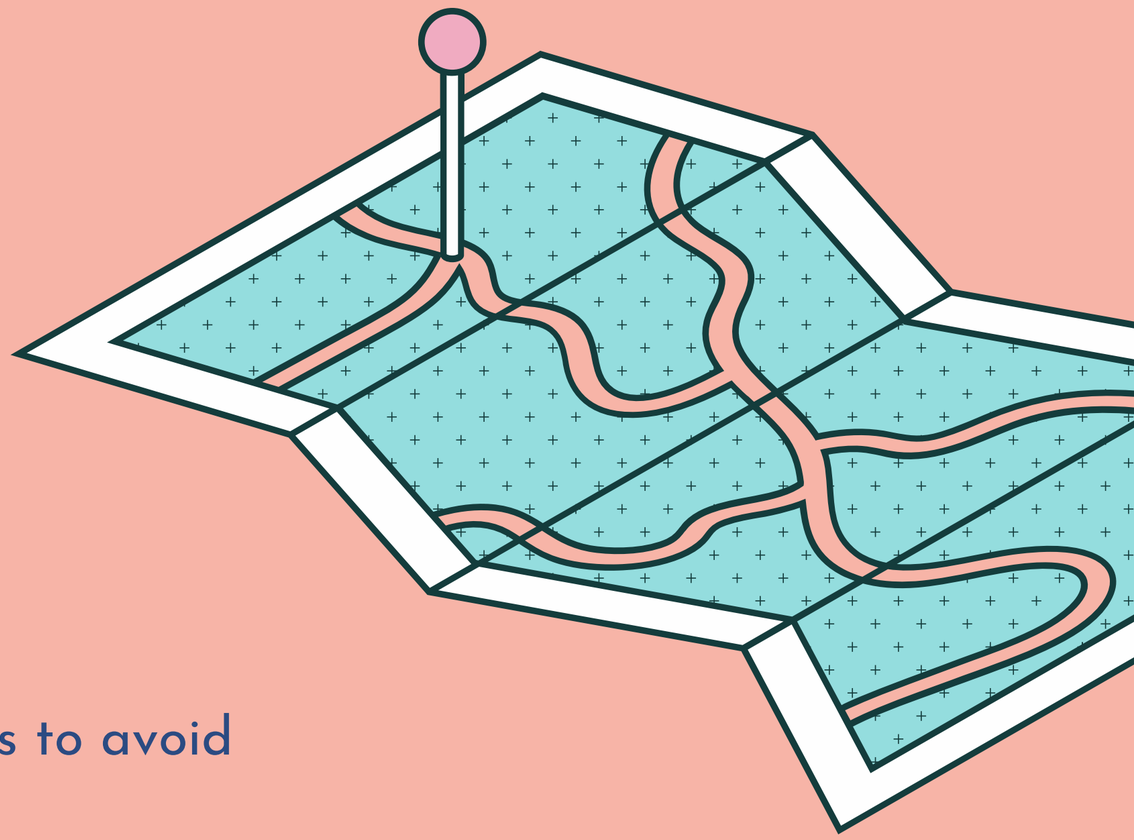
LSTM (Long Short-Term Memory Networks)

- Picks up on complex, changing trends by learning from past data.

Prophet

- Easy to use and handles holiday effects and trends without much setup.

Model Building and Training



1. Building the Model:
    - Get the data ready
    - Split it into training and test sets
2. Adjusting Model Settings:
    - Fine-tune layers, batch size, and training rounds to avoid overfitting
3. Visualizing Results:
    - Compare predicted vs. actual sales on test data
    - Look at plots to see how well the model is working

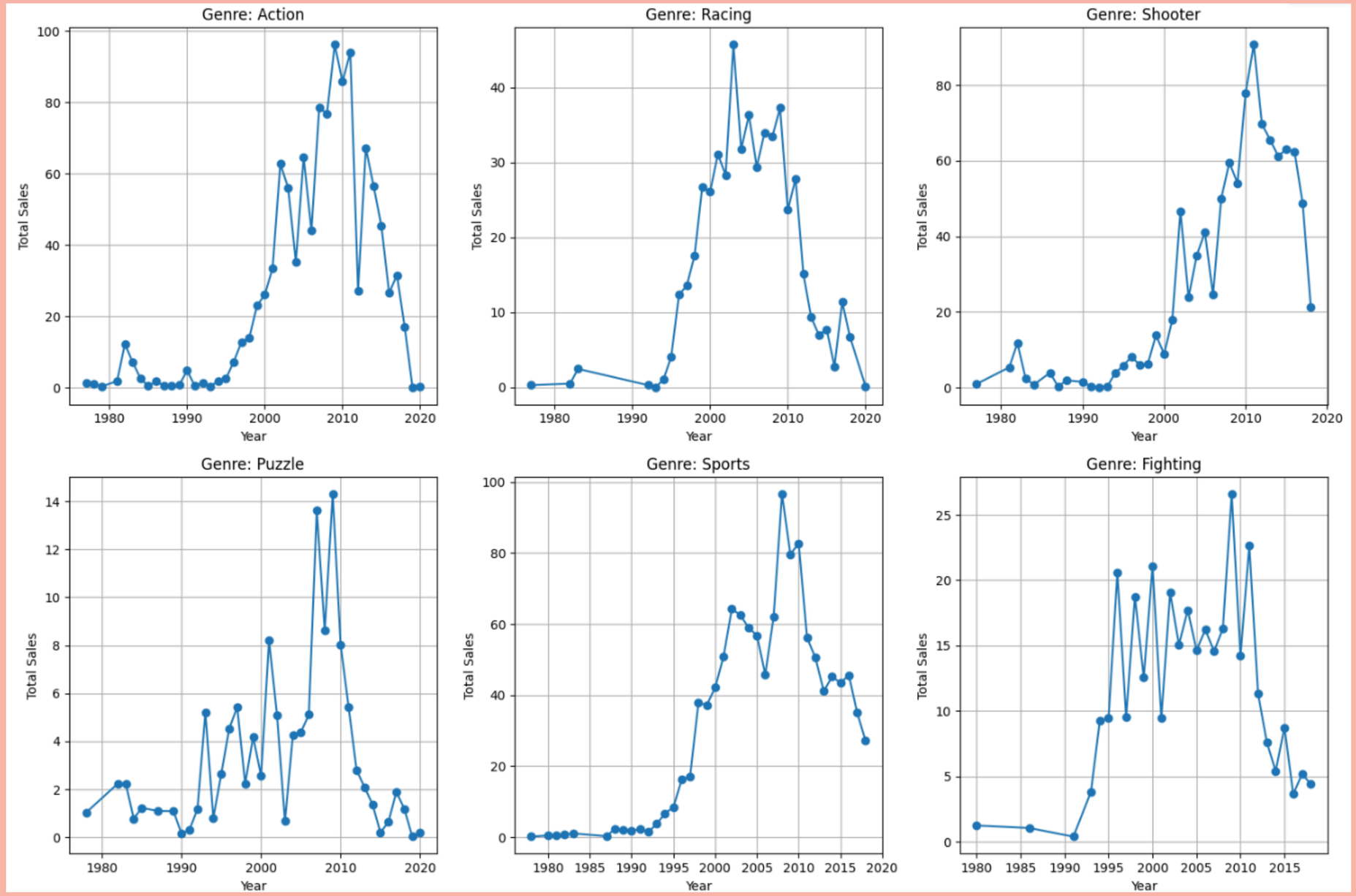
Evaluation Metrics

- **Mean Absolute Error (MAE):** Measures average error in sales predictions.
- **Mean Squared Error (MSE):** Penalizes larger errors in forecasting.

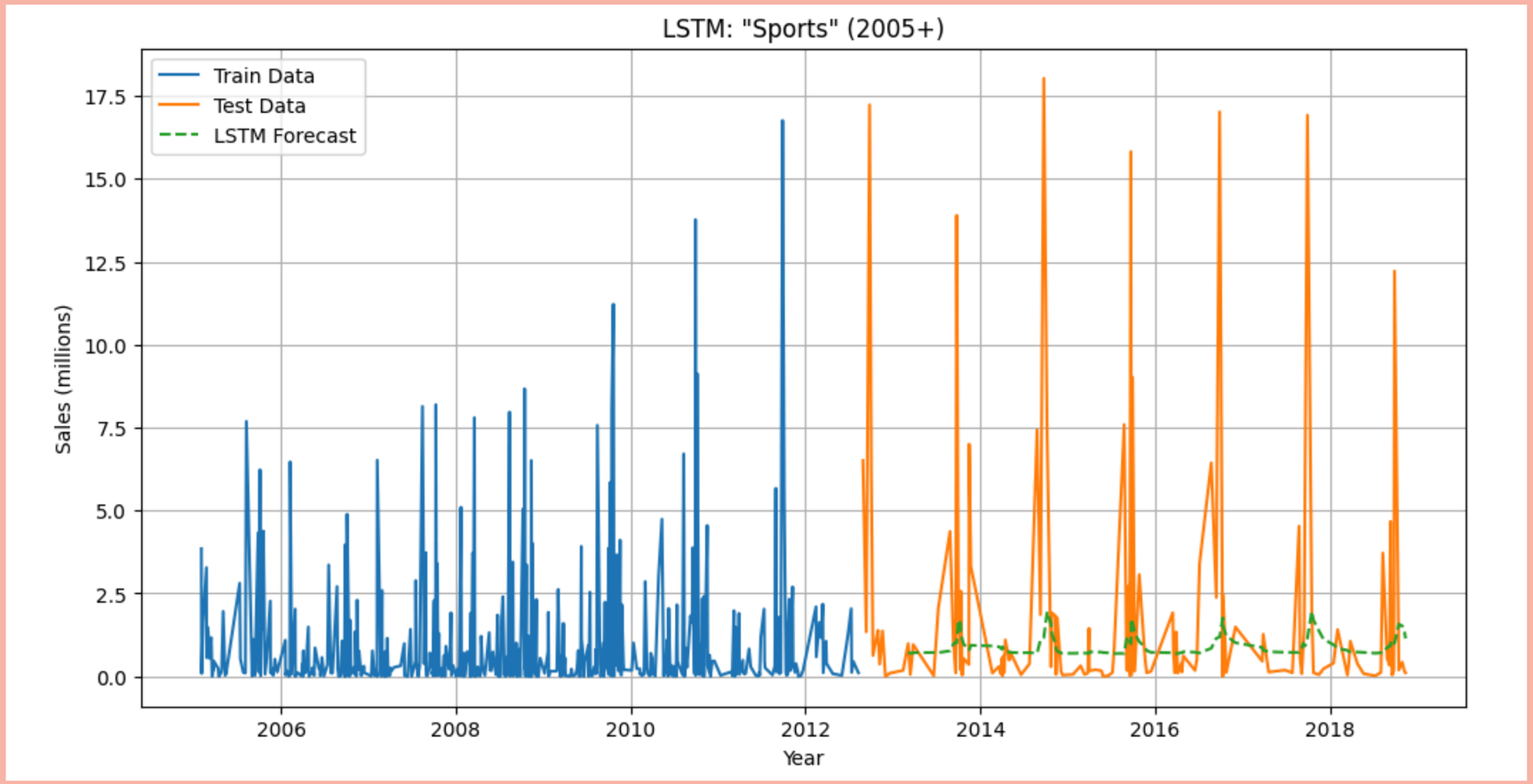
Dataset


Продажи игр по годам
Общие продажи
0
100
200
300
400
500
1980
1990
2000
2010
2020
Год

Dataset

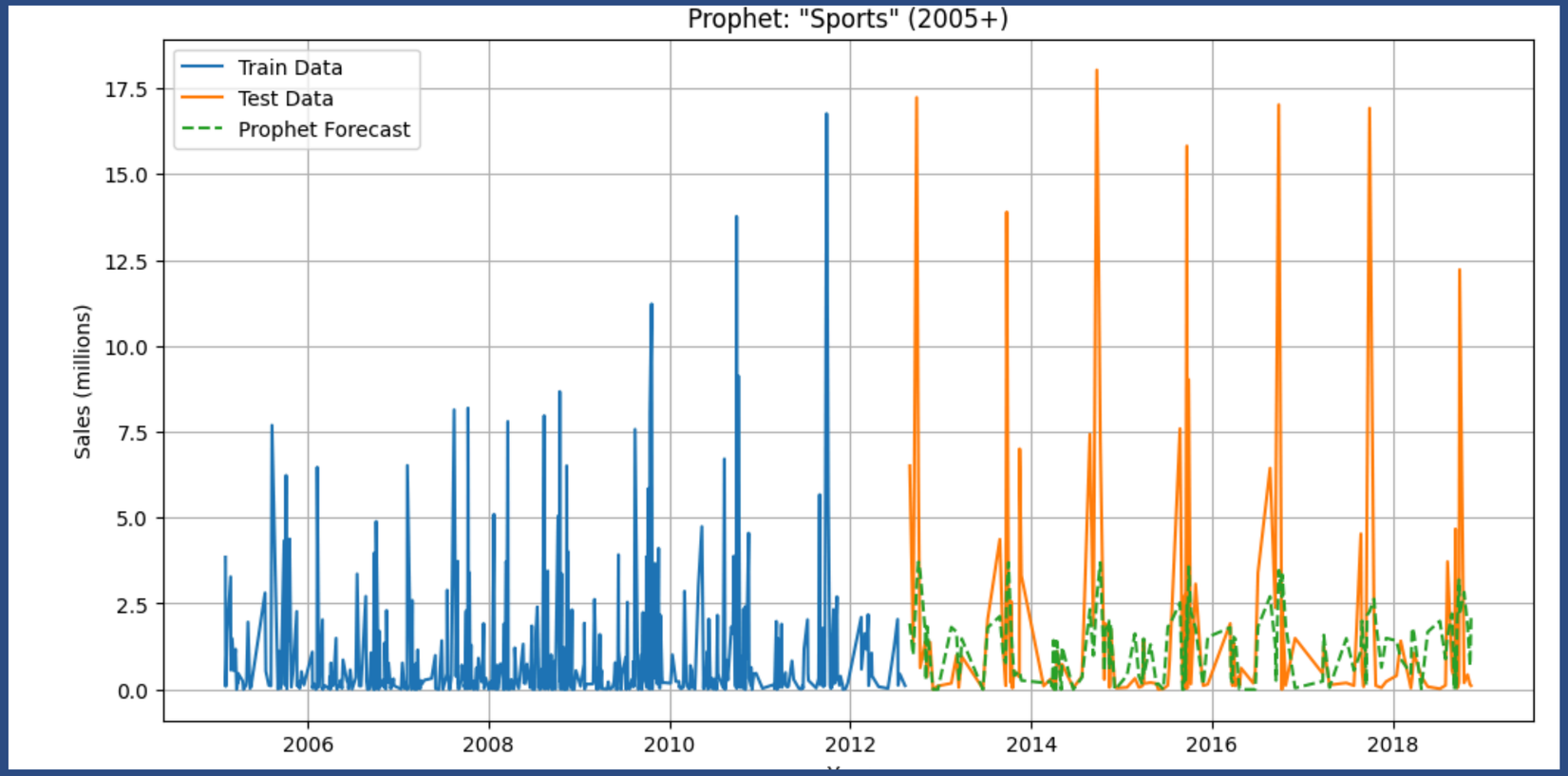

Genre: Action
Total Sales
Year
Genre: Racing
Total Sales
Year
Genre: Shooter
Total Sales
Year
Genre: Puzzle
Total Sales
Year
Genre: Sports
Total Sales
Year
Genre: Fighting
Total Sales
Year

LSTM (Long Short-Term Memory Networks)



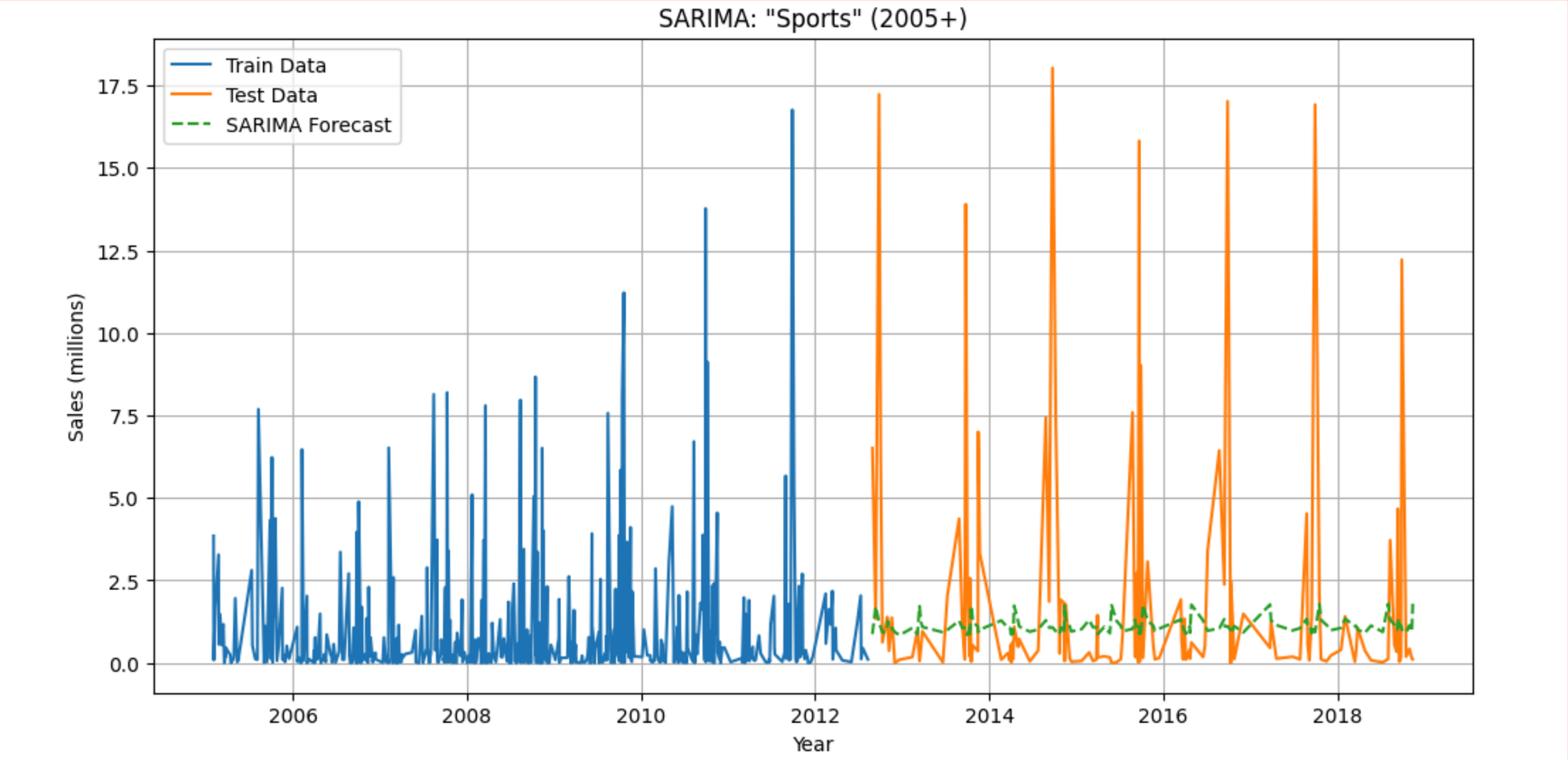
	MSE	MAE
Action	1.294148	0.945154
Shooter	26.153158	2.944014
Action-Adventure	27.932154	2.367332
Sports	14.122188	1.851039
Role-Playing	0.712330	0.462419
Simulation	0.301513	0.363126
Racing	0.586918	0.669133
Music	1.326452	1.147415
Misc	0.955921	0.574079
Fighting	0.719205	0.553873
Platform	0.817977	0.631630
Adventure	0.096260	0.262538
Strategy	0.020601	0.122261
Puzzle	0.071666	0.185843
Visual Novel	0.001106	0.030068

Prophet Algorithm



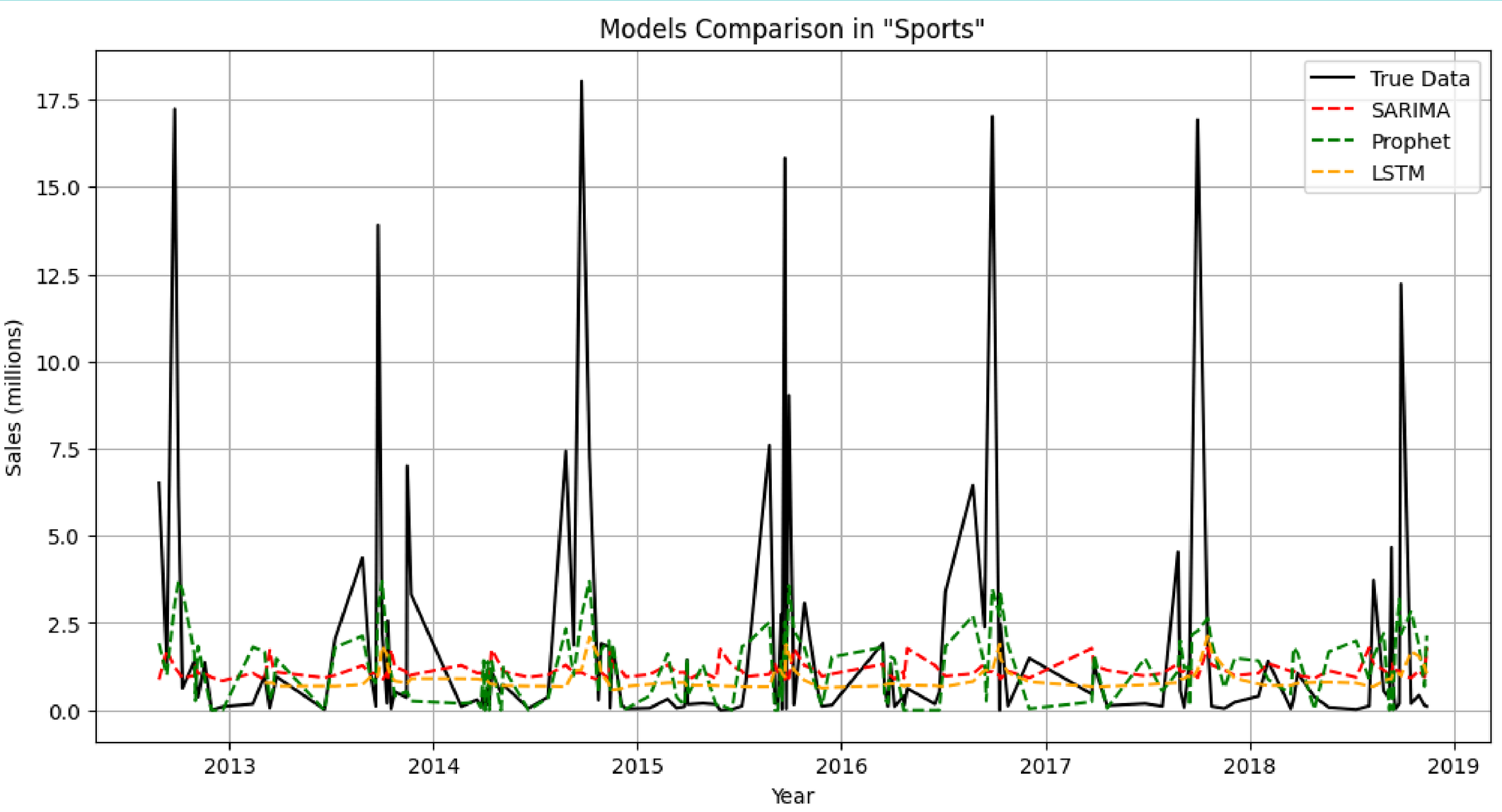
	MSE	MAE
Action	2.039668	1.077569
Shooter	19.729948	3.179852
Action-Adventure	13.170426	1.552974
Sports	12.098361	1.760546
Role-Playing	0.675062	0.466551
Simulation	0.288171	0.266375
Racing	0.760920	0.652422
Music	0.839050	0.698742
Misc	0.963613	0.426577
Fighting	0.694707	0.583917
Platform	0.857752	0.716844
Adventure	0.194706	0.275279
Strategy	0.051390	0.182301
Puzzle	0.062746	0.150774
Visual Novel	0.001553	0.028011

SARIMA (Seasonal Auto-Regressive Integrated Moving Average)



	MSE	MAE
Action	1.694458	1.106854
Shooter	24.546522	3.358277
Action-Adventure	16.835404	1.714458
Sports	15.478426	2.036738
Role-Playing	0.741123	0.581965
Simulation	0.248868	0.251096
Racing	0.833285	0.812693
Music	0.983285	0.646195
Misc	1.017680	0.385102
Fighting	0.806810	0.641629
Platform	0.980857	0.805820
Adventure	0.160339	0.208173
Strategy	0.046680	0.163134
Puzzle	0.052952	0.150419
Visual Novel	0.004969	0.048065

Comparison of three model



PROPHET

	MSE	MAE
Action	2.039668	1.077569
Shooter	19.729948	3.179852
Action-Adventure	13.170426	1.552974
Sports	12.098361	1.760546
Role-Playing	0.675062	0.466551
Simulation	0.288171	0.266375
Racing	0.760920	0.652422
Music	0.839050	0.698742
Misc	0.963613	0.426577
Fighting	0.694707	0.583917
Platform	0.857752	0.716844
Adventure	0.194706	0.275279
Strategy	0.051390	0.182301
Puzzle	0.062746	0.150774
Visual Novel	0.001553	0.028011

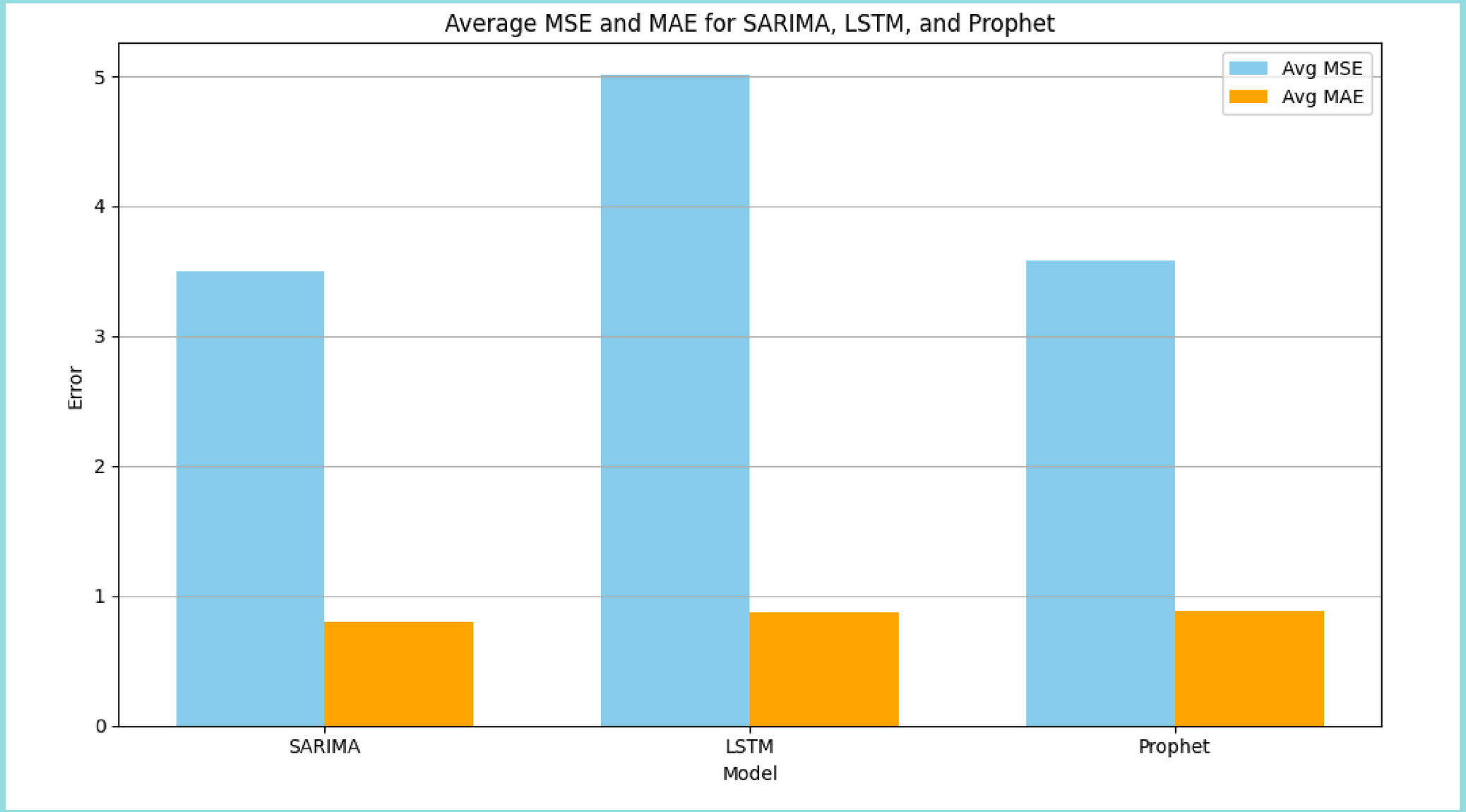
LSTM

	MSE	MAE
Action	1.294148	0.945154
Shooter	26.153158	2.944014
Action-Adventure	27.932154	2.367332
Sports	14.122188	1.851039
Role-Playing	0.712330	0.462419
Simulation	0.301513	0.363126
Racing	0.586918	0.669133
Music	1.326452	1.147415
Misc	0.955921	0.574079
Fighting	0.719205	0.553873
Platform	0.817977	0.631630
Adventure	0.096260	0.262538
Strategy	0.020601	0.122261
Puzzle	0.071666	0.185843
Visual Novel	0.001106	0.030068

SARIMA

	MSE	MAE
Action	1.694458	1.106854
Shooter	24.546522	3.358277
Action-Adventure	16.835404	1.714458
Sports	15.478426	2.036738
Role-Playing	0.741123	0.581965
Simulation	0.248868	0.251096
Racing	0.833285	0.812693
Music	0.983285	0.646195
Misc	1.017680	0.385102
Fighting	0.806810	0.641629
Platform	0.980857	0.805820
Adventure	0.160339	0.208173
Strategy	0.046680	0.163134
Puzzle	0.052952	0.150419
Visual Novel	0.004969	0.048065

Error comparison



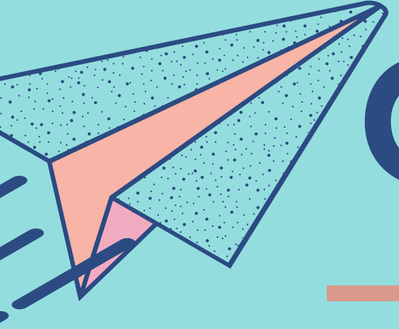
Related works

https://github.com/nktnlx/data_analysis_portfolio/blob/main/video_games_sales.ipynb

The article focuses on analyzing video game sales data from 1980 to 2016. It examines global and regional sales trends, platform popularity (e.g., PS2, Wii), temporal sales dynamics, and differences in regional preferences. Additionally, cluster analysis is used to identify groups of games with similar sales characteristics.

https://github.com/SooyeonWon/time_series_analytics/blob/main/Time_Series_Predictive_Analytics%20.ipynb

The article focuses on time series forecasting for a video game company's sales using ETS and ARIMA models. It covers data preparation, trend and seasonality analysis, model evaluation, and forecasting future sales to align supply with demand.



Conclusion

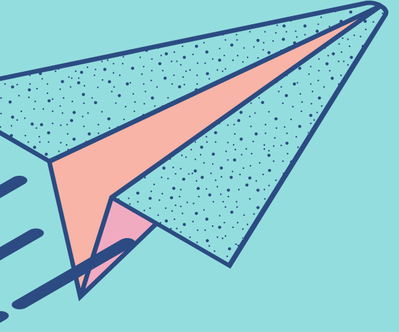
- Data collection from sources
  With focus on sales trends, genre, and platform.

- Implementing models:

The models struggled to fully capture time series patterns due to the aggregated nature of the dataset and the absence of additional features like marketing events or seasonal effects. Despite this, the chosen models were effective in handling the available data and produced reasonably low error rates.

Role of members

Timur Seitov - Coding, Preparing Datasets
Aidar Ospankulov - Coding, Contributor of Prohpet model
Olzhas Assetuly - Coding, Contributor of SARIMA model
Eldar Batkenov - Coding, Contributor of LSTM model



Thank you for your attention!