Early Stage Project Success Measurement

Presented By: Ahmed H. Alfi'er Alshareef

Table of Contents

01 Design

The problem & goal

02 Data & Algorithms

Dataset, algorithms, & architectures

03 Tools

Used, languages, frameworks & packages

04 Insights & Conclusions

Findings & prospective

01

Design

Design

- ## Why?

  1. Rapid changes in technology
  2. Entrepreneurship is the mainstream

- ## Goal:

  Determine if an idea/project is worth pursuing or not (success or fail).

Design

- ## Who?

Two main category of beneficiaries would use this project:

1. **Entrepreneurs:** To assess the quality of their idea
2. **Investors**: To determine what startups to invest in

- ## How?

Measure how likely are the users to pay for it. Success if the project achieve at least the financial goal or failure otherwise

02

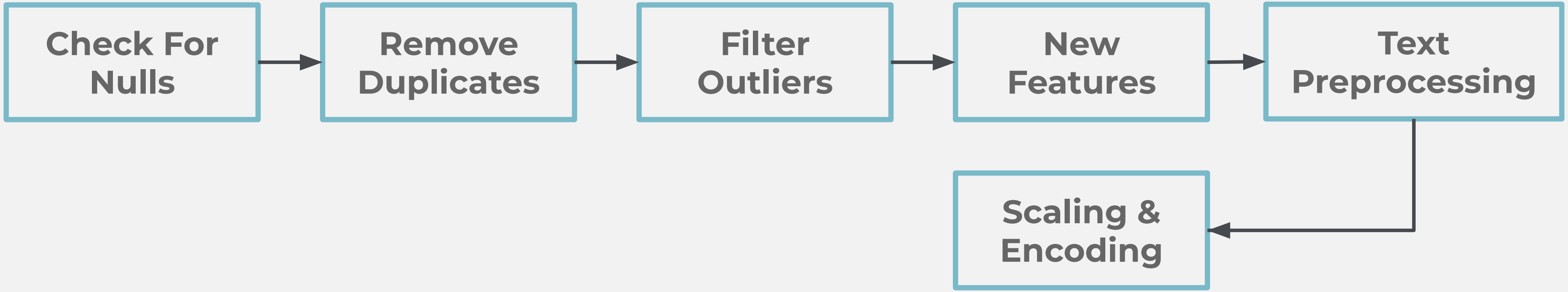
Data & Algorithms

Data & Algorithms

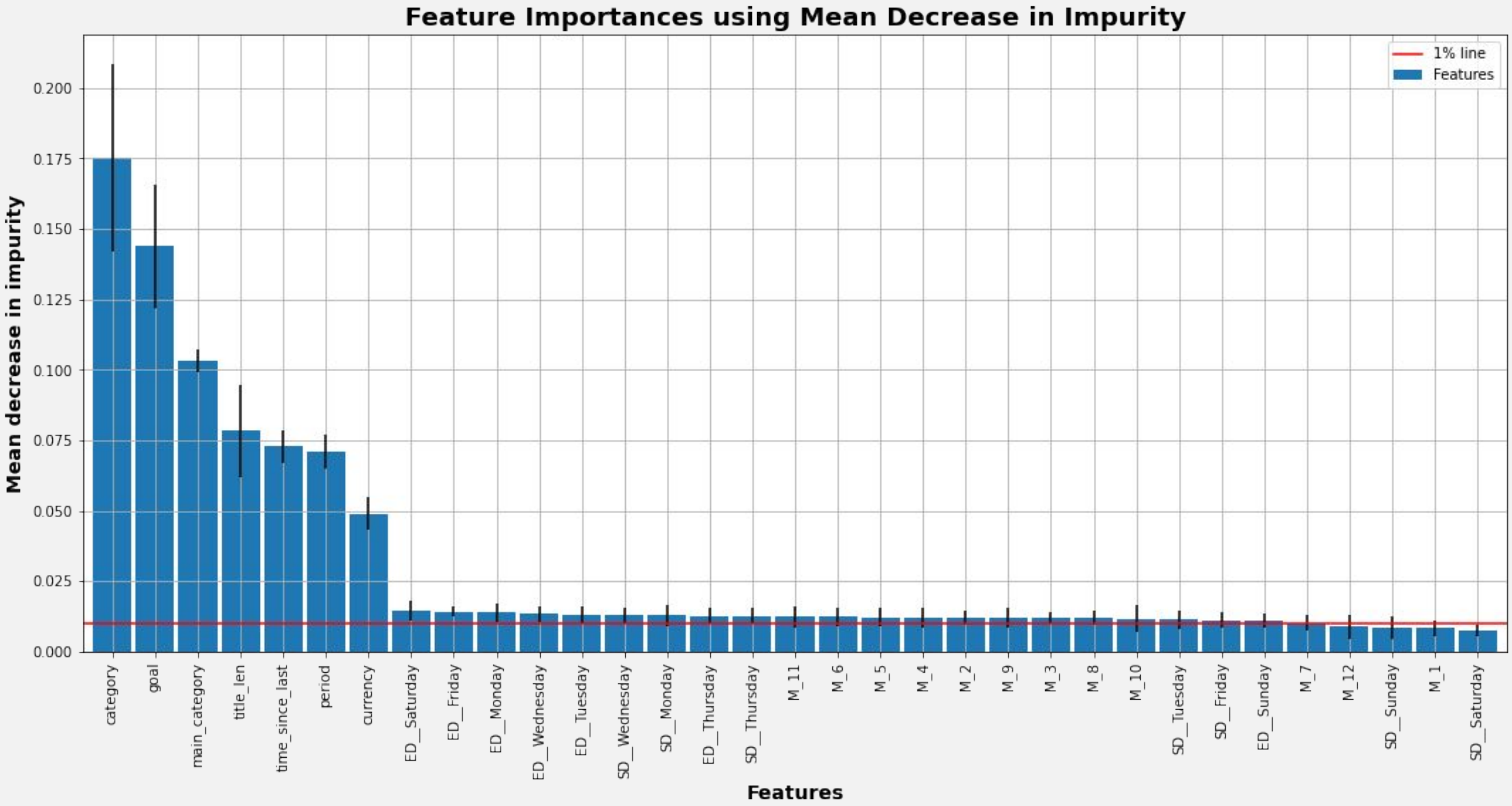
- **Data**
  1. Kickstarter dataset (Kaggel)
  2. Contains 13 columns
  3. 378,661 projects → After cleaning (124,235)

Data & Algorithms

- **Preprocessing /Features Engineering:**
  - The Preprocessing pipeline



Data & Algorithms



Feature Importances using Mean Decrease in Impurity

Data & Algorithms

- ## Algorithms:

Since it is a classification problem, several models were tested:

1. Classical Models (Logistic Regression & Support Vector Machine)
2. Ensemble **Bagging** Models (Random Forest)
3. Ensemble **Boosting** Models (Gradient Boosting)
4. Ensemble **Stacking** Models (Bert + Gradient Boosting)
5. Deep Learning Sequence Models (Bidirectional LSTM)
6. Pre-trained Models (Bert)

Data & Algorithms

Metrics	Logistic Regression	Support Vector Machine	Random Forest	Gradient Boosting	Bert + Gradient Boosting	Bi- LSTM	Bi- LSTM + NN	Bert
Accuracy	0.7445	0.7442	0.7447	**0.7512**	0.7505	0.7334	0.2666	0.7304
Precision	0.5664	0.5946	0.5539	**0.5939**	0.5905	0	0.2666	0.4838
Recall	0.1928	0.1375	0.2348	**0.2210**	0.2199	0	1	0.1647
F1	0.2877	0.2234	0.3298	**0.3221**	0.3204	-	-	-
AUC	0.7185	0.7184	0.7190	**0.7384**	0.7381	0.5	0.5	0.6218

03

Tools

Tools

1. **Data Processing:**

   Pandas, and Numpy

2. **Modelling:**

   SciKit-Learn, PyTorch, TensorFlow/Keras, and Pre-trained models (Bert & Glov)

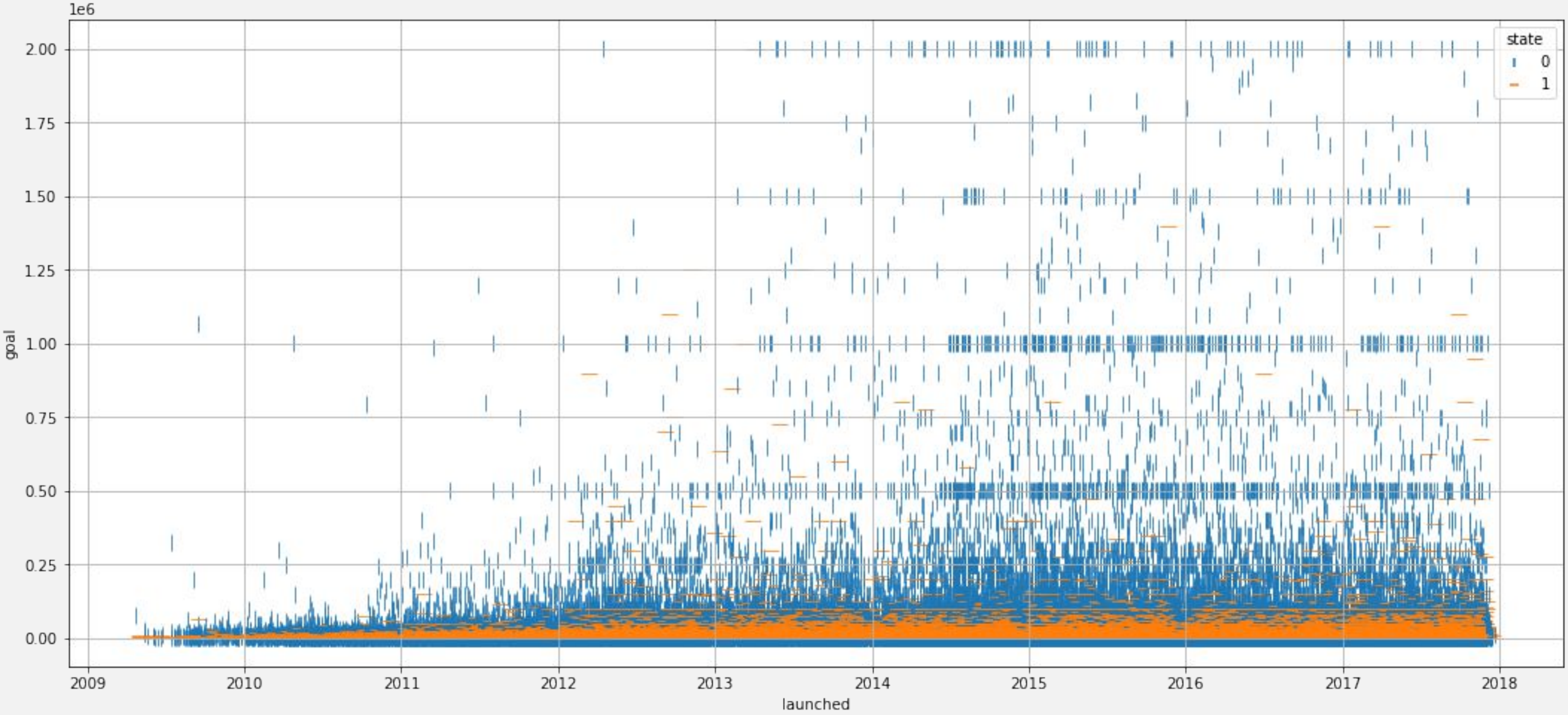
3. **Visualization:**

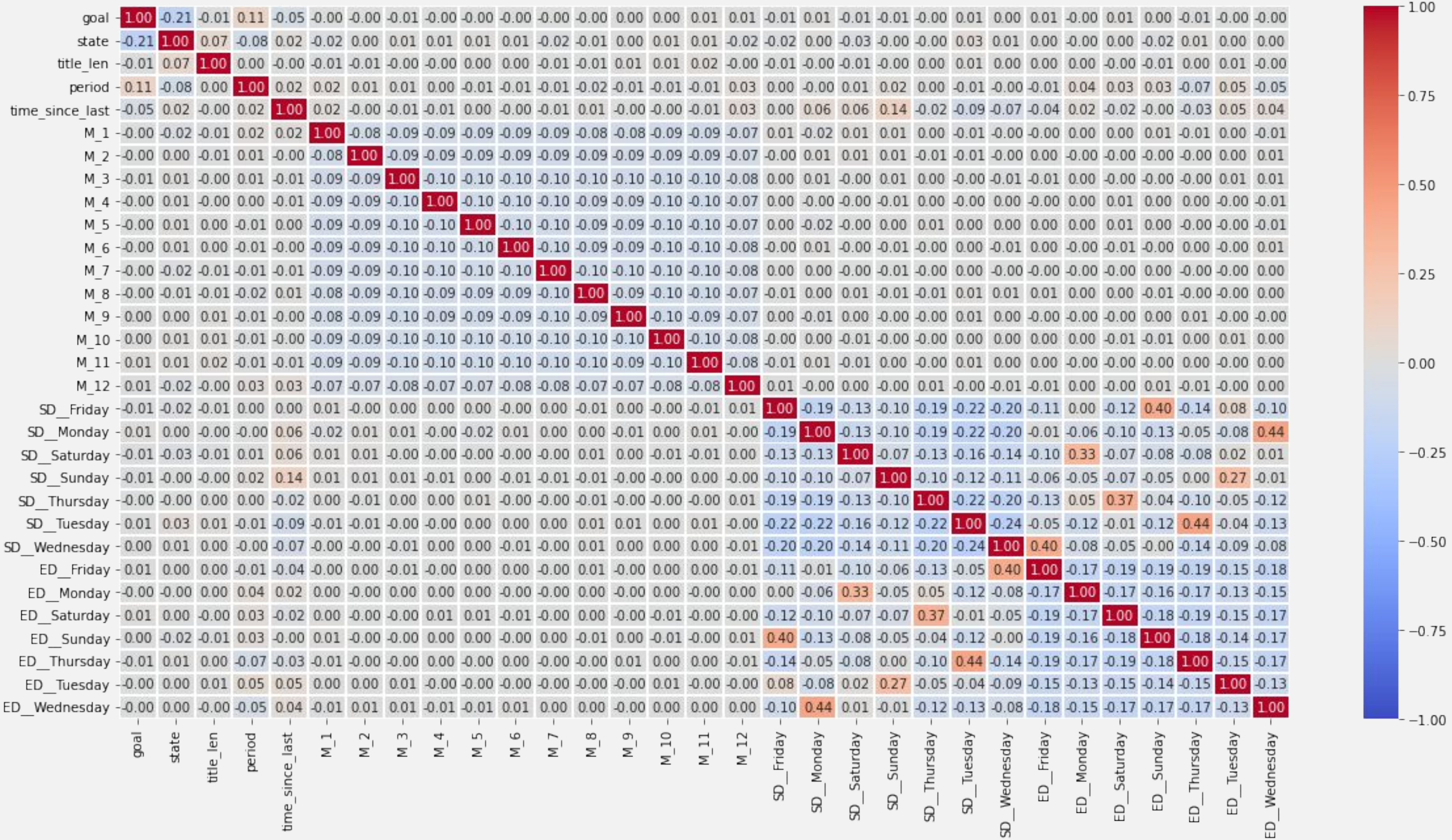
   Matplotlib, Seaborn, and Google Colab

04

Insights & Conclusion

Insights & Conclusion





	goal	state	title_len	period	time_since_last	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_10	M_11	M_12	SD__Friday	SD__Monday	SD__Saturday	SD__Sunday	SD__Thursday	SD__Tuesday	SD__Wednesday	ED__Friday	ED__Monday	ED__Saturday	ED__Sunday	ED__Thursday	ED__Tuesday	ED__Wednesday
goal	1.00	-0.21	-0.01	0.11	-0.05	-0.00	-0.00	-0.01	-0.00	-0.00	-0.00	-0.00	-0.00	0.00	0.00	0.01	0.01	-0.01	0.01	-0.01	-0.01	-0.00	0.01	0.00	0.01	-0.00	0.01	0.00	-0.01	-0.00	-0.00
state	-0.21	1.00	0.07	-0.08	0.02	-0.02	0.00	0.01	0.01	0.01	0.01	-0.02	-0.01	0.00	0.01	0.01	-0.02	-0.02	0.00	-0.03	-0.00	-0.00	0.03	0.01	0.00	-0.00	0.00	-0.02	0.01	0.00	0.00
title_len	-0.01	0.07	1.00	0.00	-0.00	-0.01	-0.01	-0.00	-0.00	0.00	0.00	-0.01	-0.01	0.01	0.01	0.02	-0.00	-0.01	-0.00	-0.01	-0.00	0.00	0.01	0.00	0.00	0.00	-0.00	-0.01	0.00	0.01	-0.00
period	0.11	-0.08	0.00	1.00	0.02	0.02	0.01	0.01	0.00	-0.01	-0.01	-0.01	-0.02	-0.01	-0.01	-0.01	0.03	0.00	-0.00	0.01	0.02	0.00	-0.01	-0.00	-0.01	0.04	0.03	0.03	-0.07	0.05	-0.05
time_since_last	-0.05	0.02	-0.00	0.02	1.00	0.02	-0.00	-0.01	-0.01	0.00	-0.00	-0.01	0.01	-0.00	-0.00	-0.01	0.03	0.00	0.06	0.06	0.14	-0.02	-0.09	-0.07	-0.04	0.02	-0.02	-0.00	-0.03	0.05	0.04
M_1	-0.00	-0.02	-0.01	0.02	0.02	1.00	-0.08	-0.09	-0.09	-0.09	-0.09	-0.09	-0.08	-0.08	-0.09	-0.09	-0.07	0.01	-0.02	0.01	0.01	0.00	-0.01	-0.00	-0.00	0.00	0.00	0.01	-0.01	0.00	-0.01
M_2	-0.00	0.00	-0.01	0.01	-0.00	-0.08	1.00	-0.09	-0.09	-0.09	-0.09	-0.09	-0.09	-0.09	-0.09	-0.09	-0.07	-0.00	0.01	0.01	0.01	-0.01	-0.01	-0.00	0.00	-0.00	-0.00	-0.00	-0.00	0.00	0.01
M_3	-0.01	0.01	-0.00	0.01	-0.01	-0.09	-0.09	1.00	-0.10	-0.10	-0.10	-0.10	-0.10	-0.10	-0.10	-0.10	-0.08	0.00	0.01	-0.00	0.01	0.00	-0.00	-0.01	-0.01	0.00	-0.00	-0.00	-0.00	0.01	0.01
M_4	-0.00	0.01	-0.00	0.00	-0.01	-0.09	-0.09	-0.10	1.00	-0.10	-0.10	-0.10	-0.09	-0.09	-0.10	-0.10	-0.07	0.00	-0.00	-0.00	-0.01	0.00	-0.00	0.00	-0.00	0.00	0.01	0.00	0.00	-0.00	-0.01
M_5	-0.00	0.01	0.00	-0.01	0.00	-0.09	-0.09	-0.10	-0.10	1.00	-0.10	-0.10	-0.09	-0.09	-0.10	-0.10	-0.07	0.00	-0.02	-0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.01	0.00	-0.00	-0.00	-0.01
M_6	-0.00	0.01	0.00	-0.01	-0.00	-0.09	-0.09	-0.10	-0.10	-0.10	1.00	-0.10	-0.09	-0.09	-0.10	-0.10	-0.08	-0.00	0.01	-0.00	-0.01	-0.00	0.00	-0.01	0.00	-0.00	-0.01	-0.00	0.00	-0.00	0.01
M_7	-0.00	-0.02	-0.01	-0.01	-0.01	-0.09	-0.09	-0.10	-0.10	-0.10	-0.10	1.00	-0.10	-0.10	-0.10	-0.10	-0.08	0.00	0.00	-0.00	-0.01	-0.00	0.00	-0.00	-0.00	-0.00	-0.00	0.00	-0.00	-0.00	0.00
M_8	-0.00	-0.01	-0.01	-0.02	0.01	-0.08	-0.09	-0.10	-0.09	-0.09	-0.09	-0.10	1.00	-0.09	-0.10	-0.10	-0.07	-0.01	0.00	0.01	-0.01	-0.01	0.01	0.01	0.01	0.00	0.00	-0.01	-0.00	-0.00	0.00
M_9	0.00	0.00	0.01	-0.01	-0.00	-0.08	-0.09	-0.10	-0.09	-0.09	-0.09	-0.10	-0.09	1.00	-0.10	-0.09	-0.07	0.00	-0.01	0.00	-0.00	-0.00	0.01	0.00	-0.00	-0.00	-0.00	0.00	0.01	-0.00	-0.00
M_10	0.00	0.01	0.01	-0.01	-0.00	-0.09	-0.09	-0.10	-0.10	-0.10	-0.10	-0.10	-0.10	-0.10	1.00	-0.10	-0.08	-0.00	0.00	-0.01	-0.00	-0.00	0.00	0.00	0.00	0.00	-0.01	-0.01	0.00	0.01	0.00
M_11	0.01	0.01	0.02	-0.01	-0.01	-0.09	-0.09	-0.10	-0.10	-0.10	-0.10	-0.10	-0.10	-0.09	-0.10	1.00	-0.08	-0.01	0.01	-0.01	0.00	-0.00	0.01	0.00	0.00	-0.00	-0.00	-0.00	0.00	-0.00	0.00
M_12	0.01	-0.02	-0.00	0.03	0.03	-0.07	-0.07	-0.08	-0.07	-0.07	-0.08	-0.08	-0.07	-0.07	-0.08	-0.08	1.00	0.01	-0.00	0.00	-0.00	0.01	-0.00	-0.01	-0.01	0.00	-0.00	0.01	-0.01	-0.00	0.00
SD__Friday	-0.01	-0.02	-0.01	0.00	0.00	0.01	-0.00	0.00	0.00	0.00	-0.00	0.00	-0.01	0.00	-0.00	-0.01	0.01	1.00	-0.19	-0.13	-0.10	-0.19	-0.22	-0.20	-0.11	0.00	-0.12	0.40	-0.14	0.08	-0.10
SD__Monday	0.01	0.00	-0.00	-0.00	0.06	-0.02	0.01	0.01	-0.00	-0.02	0.01	0.00	0.00	-0.01	0.00	0.01	-0.00	-0.19	1.00	-0.13	-0.10	-0.19	-0.22	-0.20	-0.01	-0.06	-0.10	-0.13	-0.05	-0.08	0.44
SD__Saturday	-0.01	-0.03	-0.01	0.01	0.06	0.01	0.01	-0.00	-0.00	-0.00	-0.00	-0.00	0.01	0.00	-0.01	-0.01	0.00	-0.13	-0.13	1.00	-0.07	-0.13	-0.16	-0.14	-0.10	0.33	-0.07	-0.08	-0.08	0.02	0.01
SD__Sunday	-0.01	-0.00	-0.00	0.02	0.14	0.01	0.01	0.01	-0.01	0.00	-0.01	-0.01	-0.01	-0.00	-0.00	0.00	-0.00	-0.10	-0.10	-0.07	1.00	-0.10	-0.12	-0.11	-0.06	-0.05	-0.07	-0.05	0.00	0.27	-0.01
SD__Thursday	-0.00	-0.00	0.00	0.00	-0.02	0.00	-0.01	0.00	0.00	0.01	-0.00	-0.00	-0.01	-0.00	-0.00	-0.00	0.01	-0.19	-0.19	-0.13	-0.10	1.00	-0.22	-0.20	-0.13	0.05	0.37	-0.04	-0.10	-0.05	-0.12
SD__Tuesday	0.01	0.03	0.01	-0.01	-0.09	-0.01	-0.01	-0.00	-0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.01	-0.00	-0.22	-0.22	-0.16	-0.12	-0.22	1.00	-0.24	-0.05	-0.12	-0.01	-0.12	0.44	-0.04	-0.13
SD__Wednesday	0.00	0.01	0.00	-0.00	-0.07	-0.00	-0.00	-0.01	0.00	0.00	-0.01	-0.00	0.01	0.00	0.00	0.00	-0.01	-0.20	-0.20	-0.14	-0.11	-0.20	-0.24	1.00	0.40	-0.08	-0.05	-0.00	-0.14	-0.09	-0.08
ED__Friday	0.01	0.00	0.00	-0.01	-0.04	-0.00	0.00	-0.01	-0.00	0.00	0.00	-0.00	0.01	-0.00	0.00	0.00	-0.01	-0.11	-0.01	-0.10	-0.06	-0.13	-0.05	0.40	1.00	-0.17	-0.19	-0.19	-0.19	-0.15	-0.18
ED__Monday	-0.00	-0.00	0.00	0.04	0.02	0.00	-0.00	0.00	0.00	0.00	-0.00	-0.00	0.00	-0.00	0.00	-0.00	0.00	0.00	-0.06	0.33	-0.05	0.05	-0.12	-0.08	-0.17	1.00	-0.17	-0.16	-0.17	-0.13	-0.15
ED__Saturday	0.01	0.00	-0.00	0.03	-0.02	0.00	-0.00	-0.00	0.01	0.01	-0.01	-0.00	0.00	-0.00	-0.01	-0.00	-0.00	-0.12	-0.10	-0.07	-0.07	0.37	-0.01	-0.05	-0.19	-0.17	1.00	-0.18	-0.19	-0.15	-0.17
ED__Sunday	0.00	-0.02	-0.01	0.03	-0.00	0.01	-0.00	-0.00	0.00	0.00	-0.00	0.00	-0.01	0.00	-0.01	-0.00	0.01	0.40	-0.13	-0.08	-0.05	-0.04	-0.12	-0.00	-0.19	-0.16	-0.18	1.00	-0.18	-0.14	-0.17
ED__Thursday	-0.01	0.01	0.00	-0.07	-0.03	-0.01	-0.00	-0.00	0.00	-0.00	0.00	-0.00	-0.00	0.01	0.00	0.00	-0.01	-0.14	-0.05	-0.08	0.00	-0.10	0.44	-0.14	-0.19	-0.17	-0.19	-0.18	1.00	-0.15	-0.17
ED__Tuesday	-0.00	0.00	0.01	0.05	0.05	0.00	0.00	0.01	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	0.01	-0.00	-0.00	0.08	-0.08	0.02	0.27	-0.05	-0.04	-0.09	-0.15	-0.13	-0.15	-0.14	-0.15	1.00	-0.13
ED__Wednesday	-0.00	0.00	-0.00	-0.05	0.04	-0.01	0.01	0.01	-0.01	-0.01	0.01	0.00	0.00	-0.00	0.00	0.00	0.00	-0.10	0.44	0.01	-0.01	-0.12	-0.13	-0.08	-0.18	-0.15	-0.17	-0.17	-0.17	-0.13	1.00

Insights & Conclusion

Insights:

1. **Model Range of Prediction:** (5,000 ≤ Goal ≤ 2,000,000)
2. **Best Dates:**
   - (Launch day: Tuesday)
   - (Launch month: October)
   - (Deadline day: Thursday)
3. **Best Categories:**
   - Music
   - Theater
4. **Worst Categories:**
   - Technology
   - Food
   - Film & Video

Insights & Conclusion

Prospective:

1. **Data is not sufficient:**
   - Bias models → more complex which needs more features
   - Project description/Images
   - Unifying the currency of goal
2. **Web presence:**
   - Integrated API / Stand alone website
3. **Utilizing more GPUs & RAMs:**
   - Investigate more transformers/Pre-trained models

Thank You,

Any Questions?