Predictive Power: Unveiling Kickstarter Success

Project Team

Ilyes Azzaz

Nizar Marzougui

Grigoria Sourri

With the Guidance of Mentor Sarah Le Net

March Continuous DA 2023

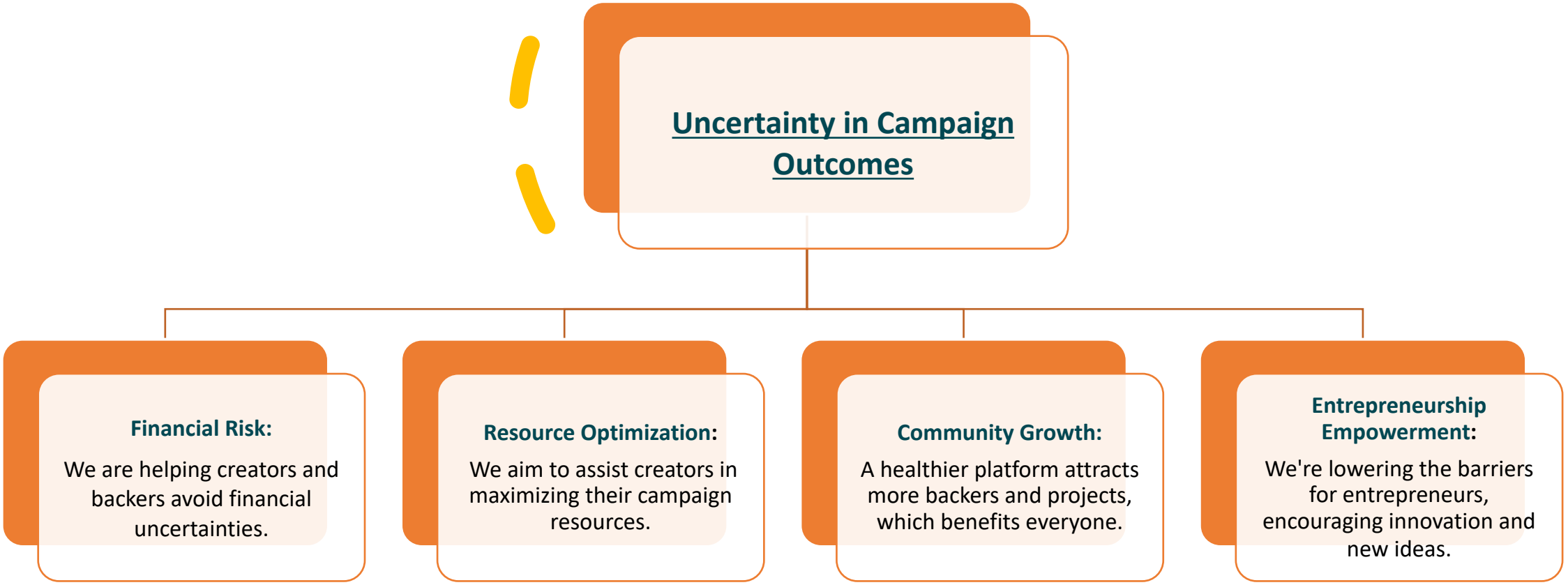
Project Objective

- Empower creators by understanding Kickstarter success
- Explore success factors and quantify variable impact
- Provide valuable insights for project success on Kickstarter

Key Questions

- What drives Kickstarter project success?
- How do project features influence success?
- What factors ensure funding goal achievement?

Addressing the Challenge and Why It Matters



Data Retrieval and Methodology

Dataset: Kickstarter campaign data

Source: Web Robots (April 2009 to August 2023)

Volume: 241392 projects campaigns

Data Architecture: Structured in CSV format

Dataset

EDA:

- Prepossessing Data
- Data Visualisation

Train and Test

Predict Model

Deploy Model

62%
of completed projects are successfully funded

How do successful and failed projects differ ?

Exploring the Factors of Success

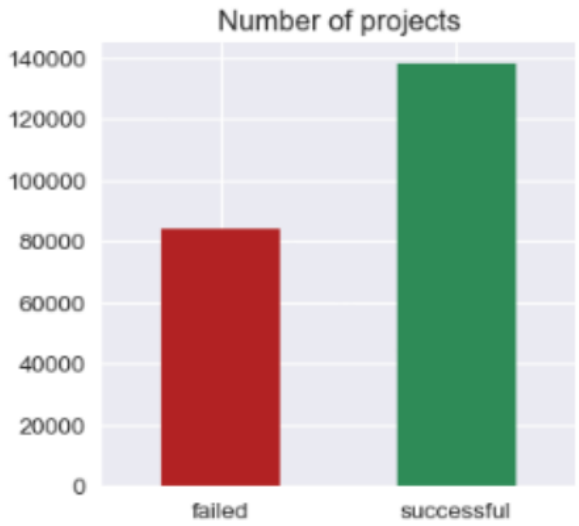
Significant portion of projects tend to meet their goals

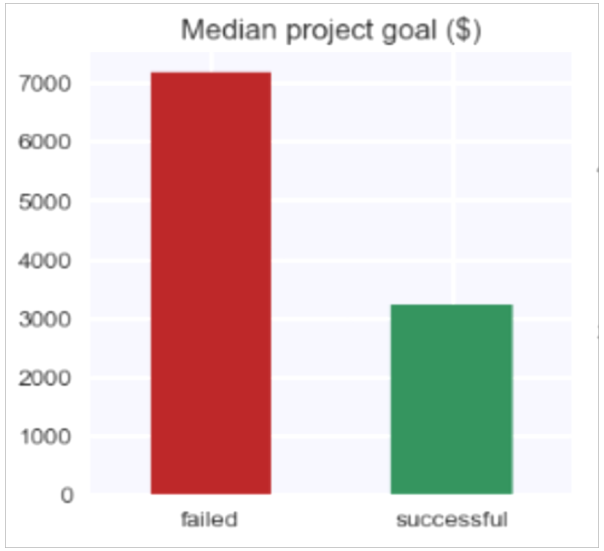
Significant portion of projects tend to meet their goals

Shorter campaign duration

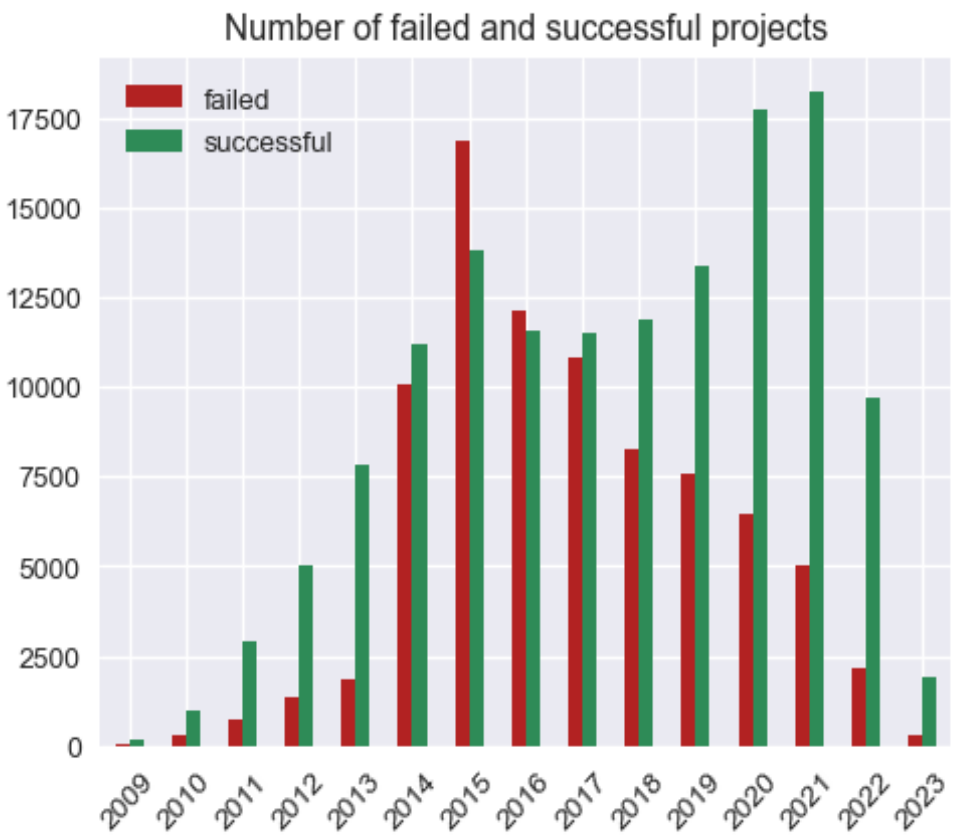
Projects selected as staff pick

- Visibility
- Credibility
- Trust



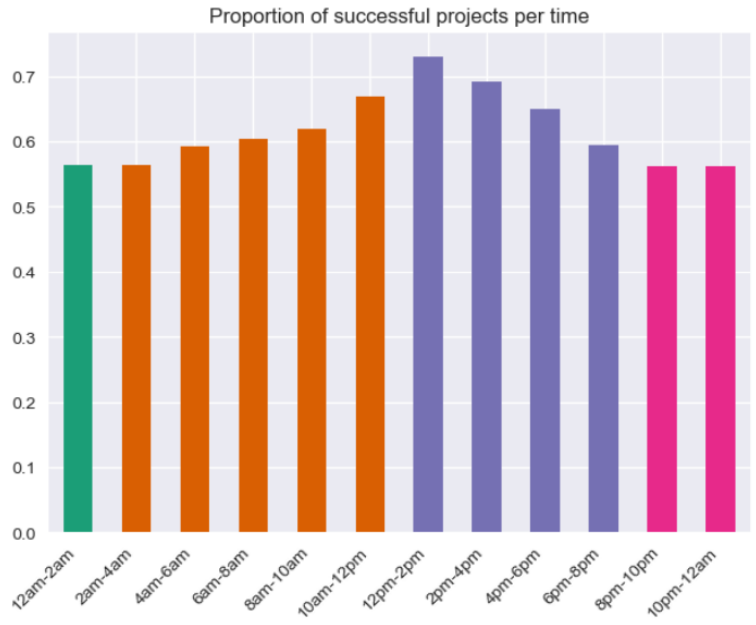


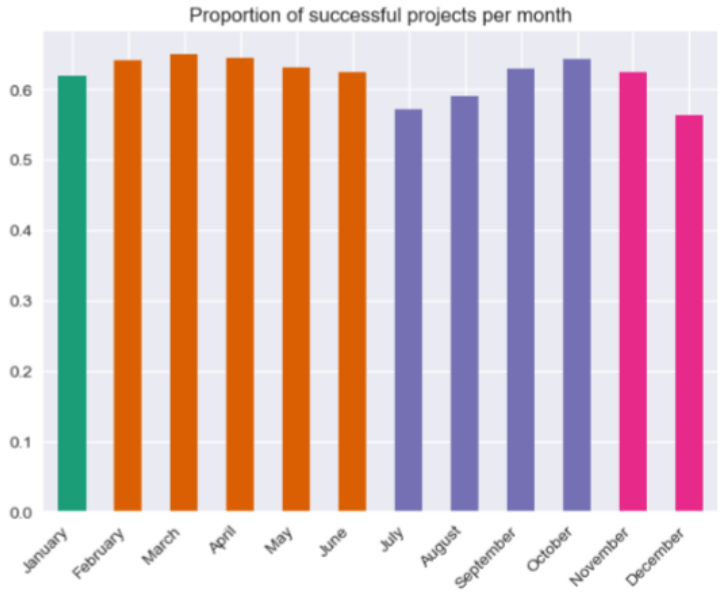
What Does the Proportion of Failed and Successful Projects Reveal?

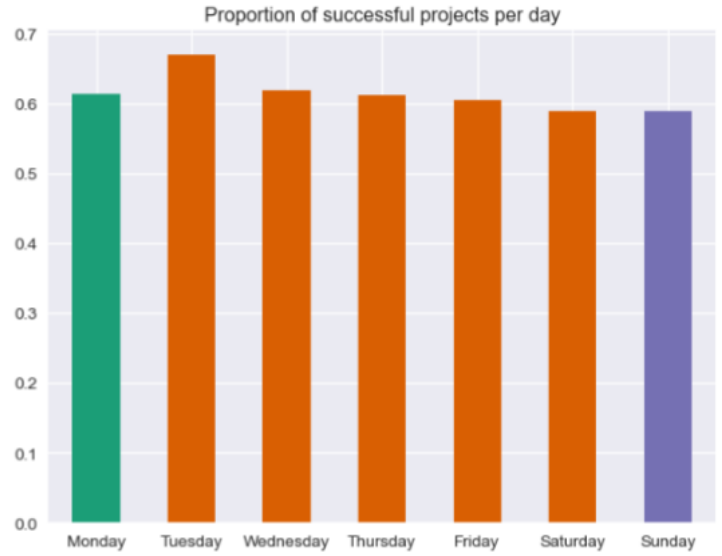


What is the best time to launch a project?

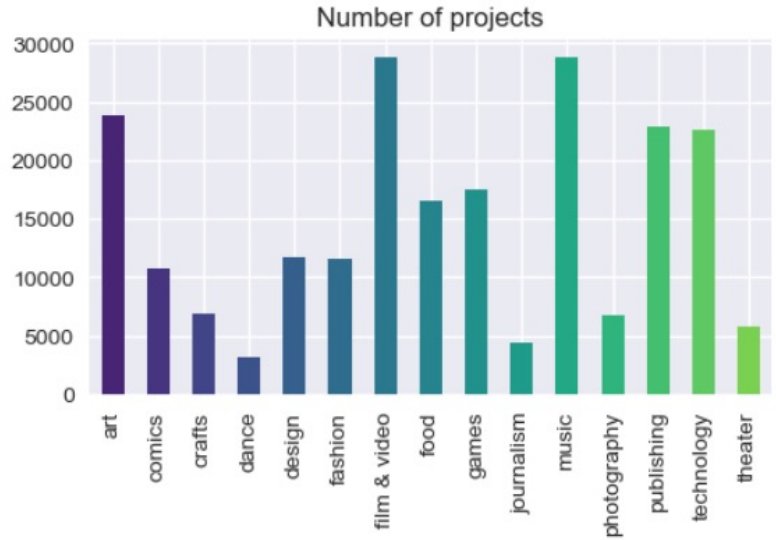
“Optimal project launch: Tuesday, 12-2 pm UTC, March”

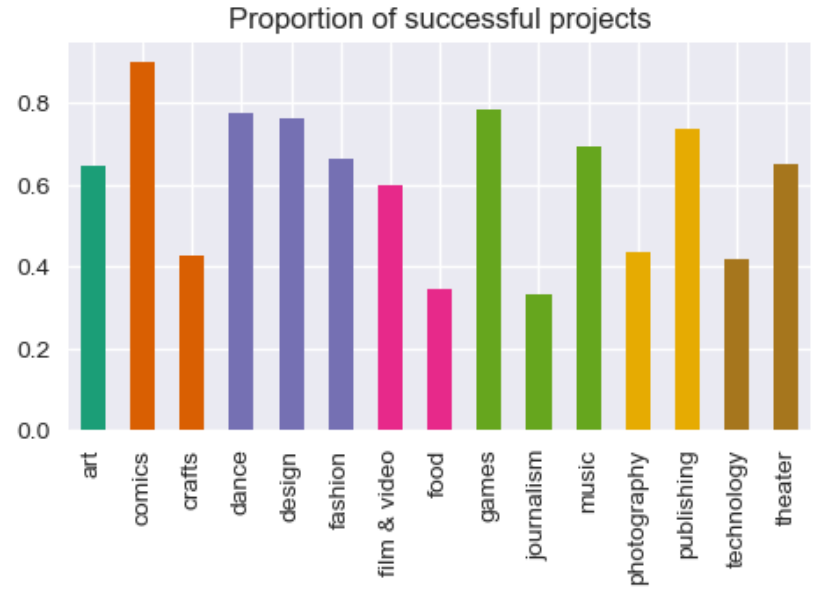






What are the Project Categories and Their Success Rates?

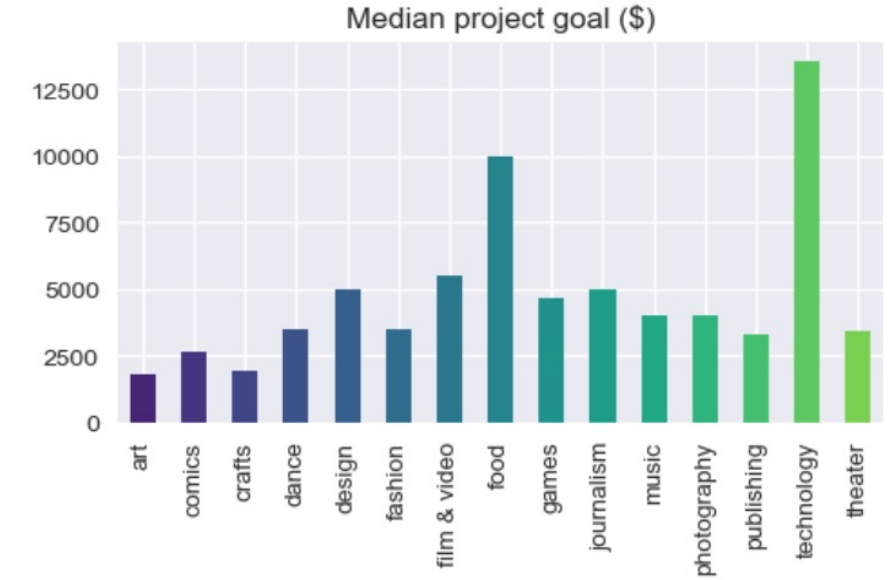




Music, **Film** & **Video**, **Art** have a high number of projects indicating **Popularity** and **High competition**

Technology and **Food** have higher median goals but relatively lower success proportions

Comics and **Dance** categories achieve more **success** due to their **modest funding goals**

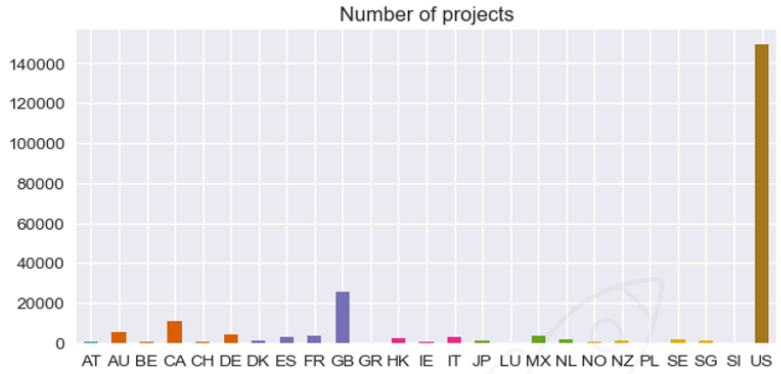


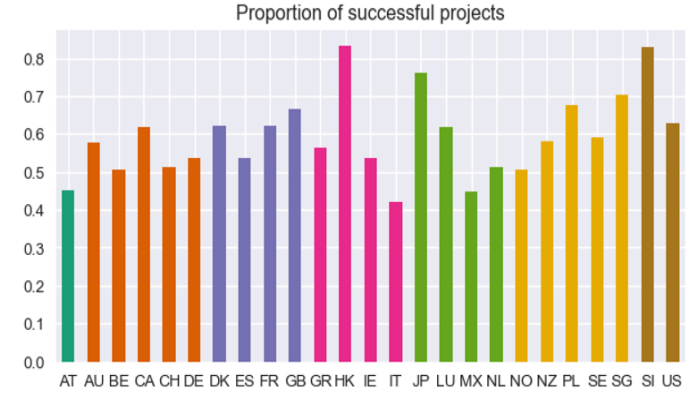
What are the Geographic Origins of Projects and Backers?

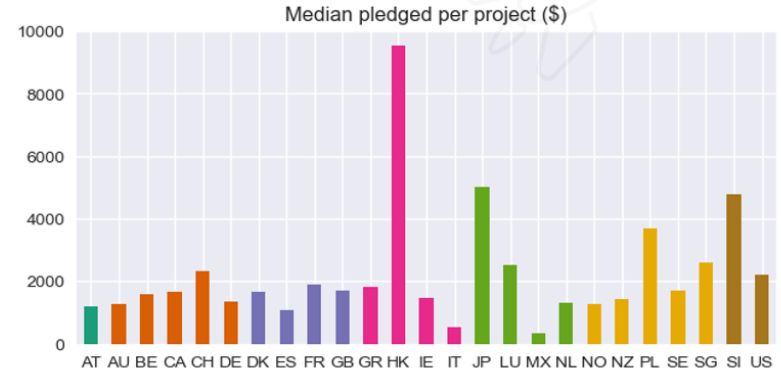
United States leads in project launches.

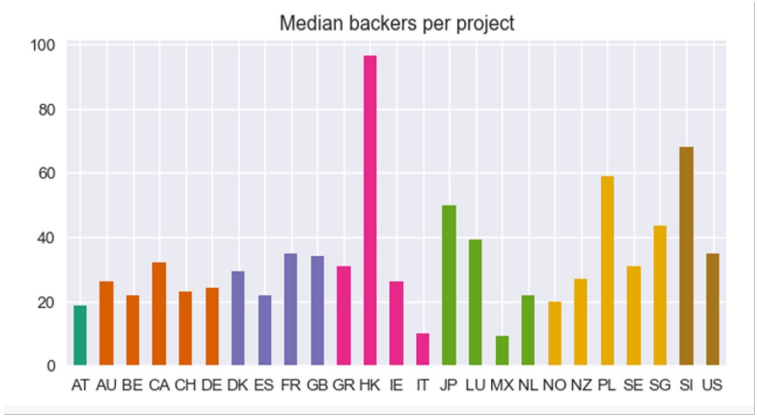
Switzerland shows the highest median project goal size.

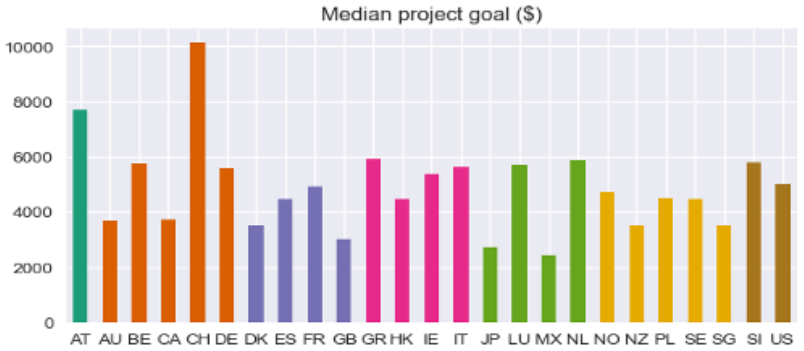
Hong Kong stands out with exceptional project success, leading in backers, funding per backer, and total funding.







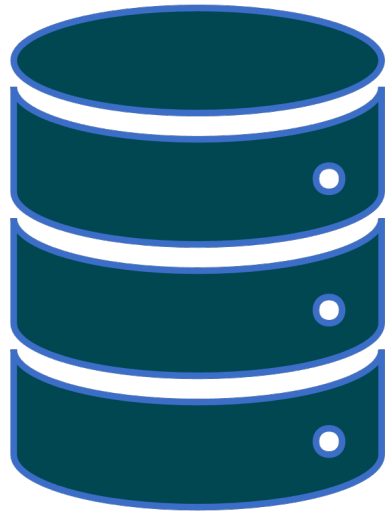




Data Preparation Pipeline

Enhancing the dataset for Machine Learning Modelling

- **Target Variable Transformation**: Converting the target variable into binary values (1 for successful projects, 0 for failed projects)
- **Multicollinearity matrix and Computing the VIF** (Variance Inflation Factor)
- **Categorical Variable Encoding.**
- **Standardization the Data .**
- **Dataset Splitting**: 80% - 20%



The Selected Models along with their Performance

1. Logistic Regression
2. Random Forest
3. Decision Tree
4. XGBoost
5. KNN

The Models along with their corresponding Training and Testing Scores:

	Model	Train Accuracy	Test Accuracy
0	Logistic Regression	0.741151	0.741851
1	Random Forest	0.999989	0.770852
2	Decision Tree	0.999994	0.689265
3	K-Nearest Neighbors	0.774257	0.657097
4	XGBoost	0.783242	0.776670

- **Overfitting:** XGBoot and Logictic Regession are the winner
- **Performance on Test Data**: XGBoost with test accuracy 77.67% is the winner.
- **Simplicity and Interpretability:** test accuracy between XGboost and Logistic Regression is not substantial. But Logistic Regressing is the winner for its simplicity.

Metrics of the Models

Data Table for Model Metric Comparison:

Model	MAE train	MSE train	RMSE train	MAE test	MSE test	RMSE test
Logistic Regression	0.258849	0.258849	0.508772	0.258149	0.258149	0.508083
Random Forest	0.116320	0.022302	0.149338	0.315511	0.158422	0.398023
Decision Tree	0.000006	0.000003	0.001676	0.311027	0.311027	0.557698
XGBoost	0.297603	0.140761	0.375181	0.310710	0.152732	0.390810
KNN	0.315264	0.152933	0.391067	0.396262	0.229103	0.478647

Comparing Model Performance: Logistic Regression, Random Forest, Decision Tree, XGBoost, and KNN"

- Best Model Based on Test Error:

Random Forest and XGBoost ==> low test error metrics

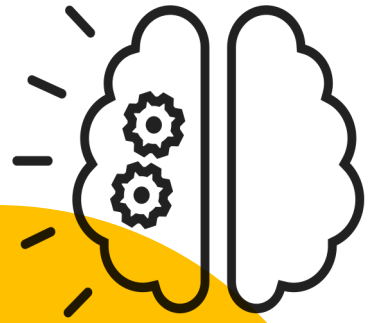
- Overfitting:

Decision Tree values very close to 0 and higher test errors, indicating an overfit.

- Consistency:

Logistic Regression and XGBoost are relatively consistent between training and test datasets.

	precision	recall	f1-score	support
0	0.75	0.61	0.68	16947
1	0.79	0.88	0.83	27570
accuracy			0.78	44517
macro avg	0.77	0.75	0.75	44517
weighted avg	0.77	0.78	0.77	44517



Model With Best Performance

XGBoost

Balance between training and test accuracy

High test accuracy

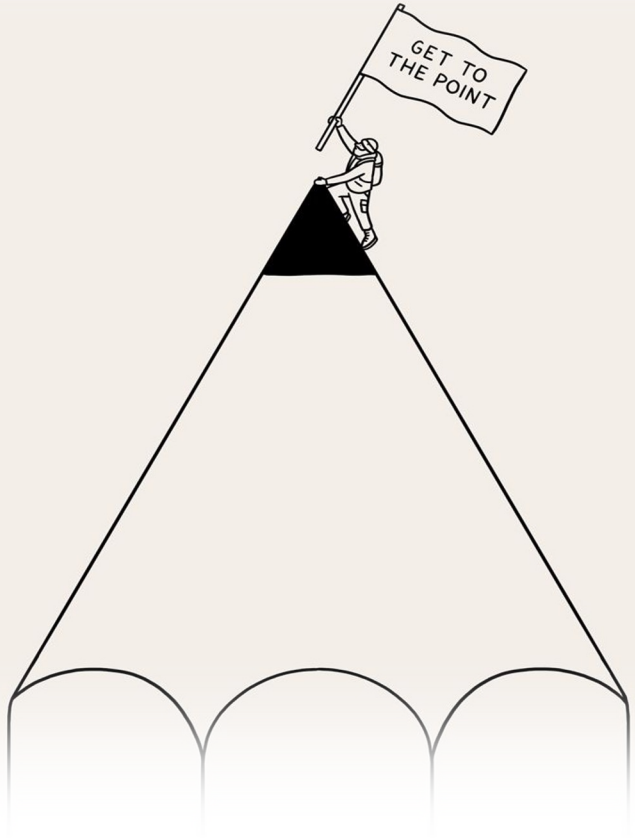
Consistent performance

Ability to overcome overfitting

Key Strategies for Crowdfunding Triumph

- ✓ **Smaller Goals:** Opt for modest funding goals for higher success.
- ✓ **Quality Matters:** Prioritize high-quality projects to attract staff picks and perform well.
- ✓ **Short Campaigns:** Keep it short, around 30 days, for better performance.
- ✓ **Pre-launch Planning:** Allocate time for sufficient preparation.
- ✓ **Explore High Success Categories:** Consider comics, dance, and games.
- ✓ **Geographic Impact:** Pay attention to project location for success and search for generous backer support.
- ✓ **Strategic Launch:** Consider launching on a Tuesday for positive results.
- ✓ **Ideal Months:** Optimal months for launch are March, April, and October, known for higher success rates.
- ✓ **Launch Timing:** Aim for a launch between 12pm and 2pm UTC, catering to the global Kickstarter audience.
- ✓ **Project Names and Blurbs:** Create shorter, compelling blurbs

Stay true to your vision



Problem addressed

- ✓ Uncertainty in Kickstarter campaign outcomes.
- ✓ Our Predictive Model enhances decision-making for creators and backers.
- ✓ The impact extends to economic benefits, resource optimization, and fostering innovation.

The Bigger Picture

- ✓ Our work benefits creators, backers, and the Kickstarter platform.
- ✓ It empowers entrepreneurs, lowers entry barriers, and contributes to platform sustainability.

Thank you for your time!