



IBM X DATASOC MACHINE LEARNING PROJECT CYCLE

OPTIMISING BUILDING PROJECT INVESTMENTS

OUR TEAM

Aaron Wong

Akhil
Panwar

Fiza
Shafeeqe

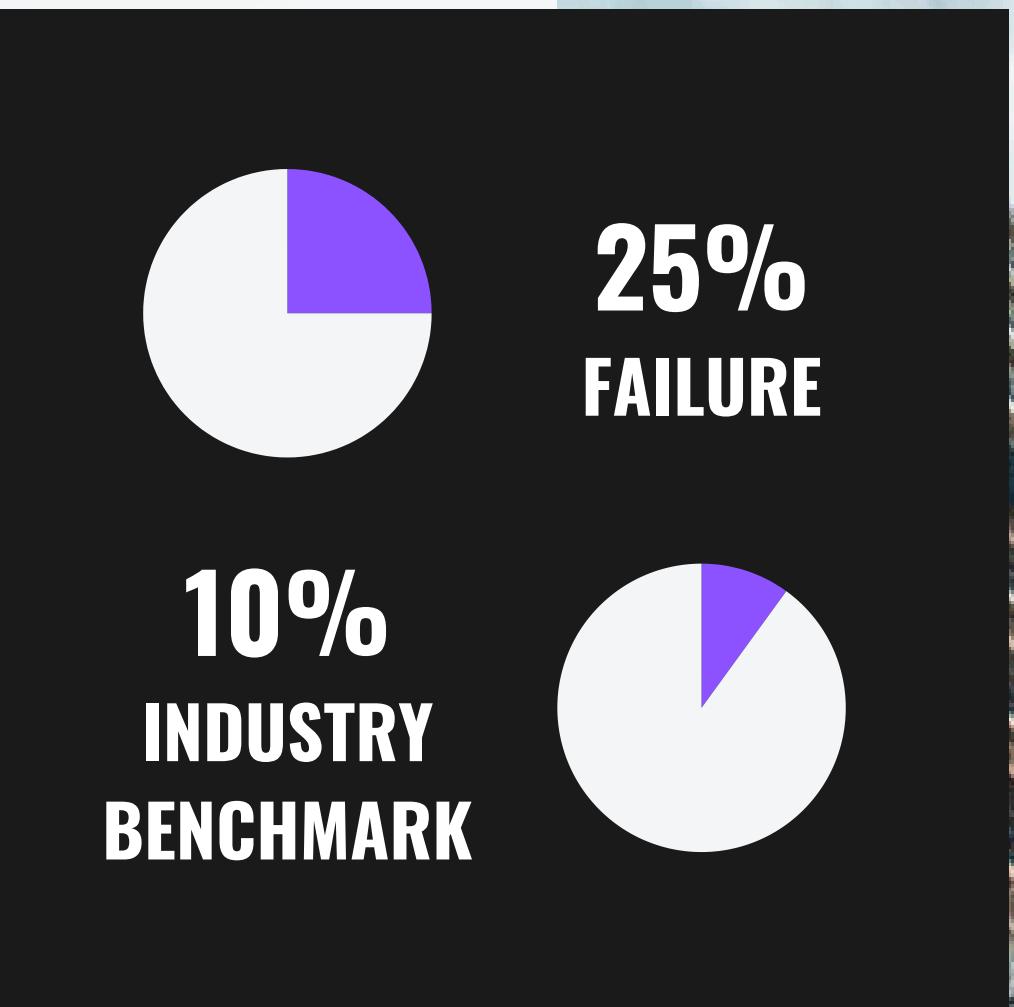
Negar
Mahdavi

Ruo Xu Shen

INTRODUCTION

Our client is a real estate developer which is currently experiencing losses for 25% of their project.

We intend to identify causes and enable the client to strategically invest in high-quality projects to maximise their profits, proactively manage risks, and enhance their project planning efficiency.



GOALS AND OBJECTIVES

Objective #1

Find what factors causes failure of a project and control these factors when planning projects for maximum profit.

Objective #2

Create a framework to search local building policies to find answers as the company plans their developments.



Preliminary Research

1.1. Poor Planning

One of the most significant factors contributing to construction project failure is poor planning. Lack of clear project goals and objectives, inadequate project scope definition, and insufficient risk management plans are all common planning-related issues that can lead to project failure. Proper planning is critical to project success as it helps ensure that everyone involved in the project understands the project goals, scope, timeline, and budget.

1.2. Ineffective Communication

Poor communication between project stakeholders is another significant cause of construction project failure. Insufficient project documentation, communication breakdowns between stakeholders, and poor communication with stakeholders outside the project team are all common communication-related issues. Effective communication is essential in construction projects, as it helps to ensure that everyone involved in the project is aware of what is happening and can make informed decisions.

1.3. Inadequate Resource Allocation

Inadequate resource allocation, including insufficient budget allocation, inadequate staffing, workforce management, and inadequate allocation of time and materials, can lead to project failure. Proper resource allocation is crucial to ensure that the project has the necessary resources to complete the project successfully.

1.4. Unforeseen Challenges

Environmental and weather-related challenges, regulatory and legal challenges, and technological and equipment failures are all examples of unforeseen challenges that can cause construction project failure. These challenges can be difficult to predict or prepare for and can have a significant impact on the project timeline and budget.

1.6. Poor Budget and Schedule Management

Poor budgets and project delays are common causes of construction project failure. Schedule delays and cost overruns can significantly impact project success, resulting in a loss of time and money. Effective budget and schedule management are crucial to ensure that the project is completed on time and within budget.

1.7. Inadequate Human Resource Management

Inadequate human resource management, including staffing and workforce management, can result in construction project failure. Proper human resource management is crucial to ensure that the project has the necessary workforce to complete the project on time and within budget.

- Poor schedule management
- Budgeting

INITIAL CONCERNS

- Majority of projects had problems with start_dates and end_dates columns.

51%

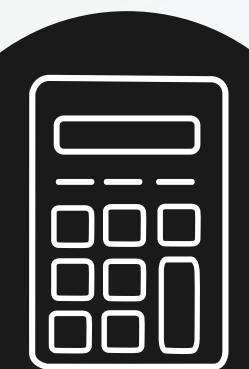
PNS	PNS	PNS
PNS	PNS	PNS
PNS	PNS	PNS
04/10/2017	01/15/2018	07/06/2017
07/18/2017	DOER	
05/24/2017	09/08/2017	08/08/2017
PNS	PNS	PNS
02/21/2017	DOER	

DATA CLEANING



- New columns for:
- Budget variance
 - Duration.

FEATURE MODIFICATION



- Convert dates to datetime format
- Encode object data type to numbers for each of quantitative analysis

DATA TYPES



- Duration: Replace with median/mean values of duration for that geographic location/project type
- Drop missing values for independent variables like spending and project type

MISSING VALUES

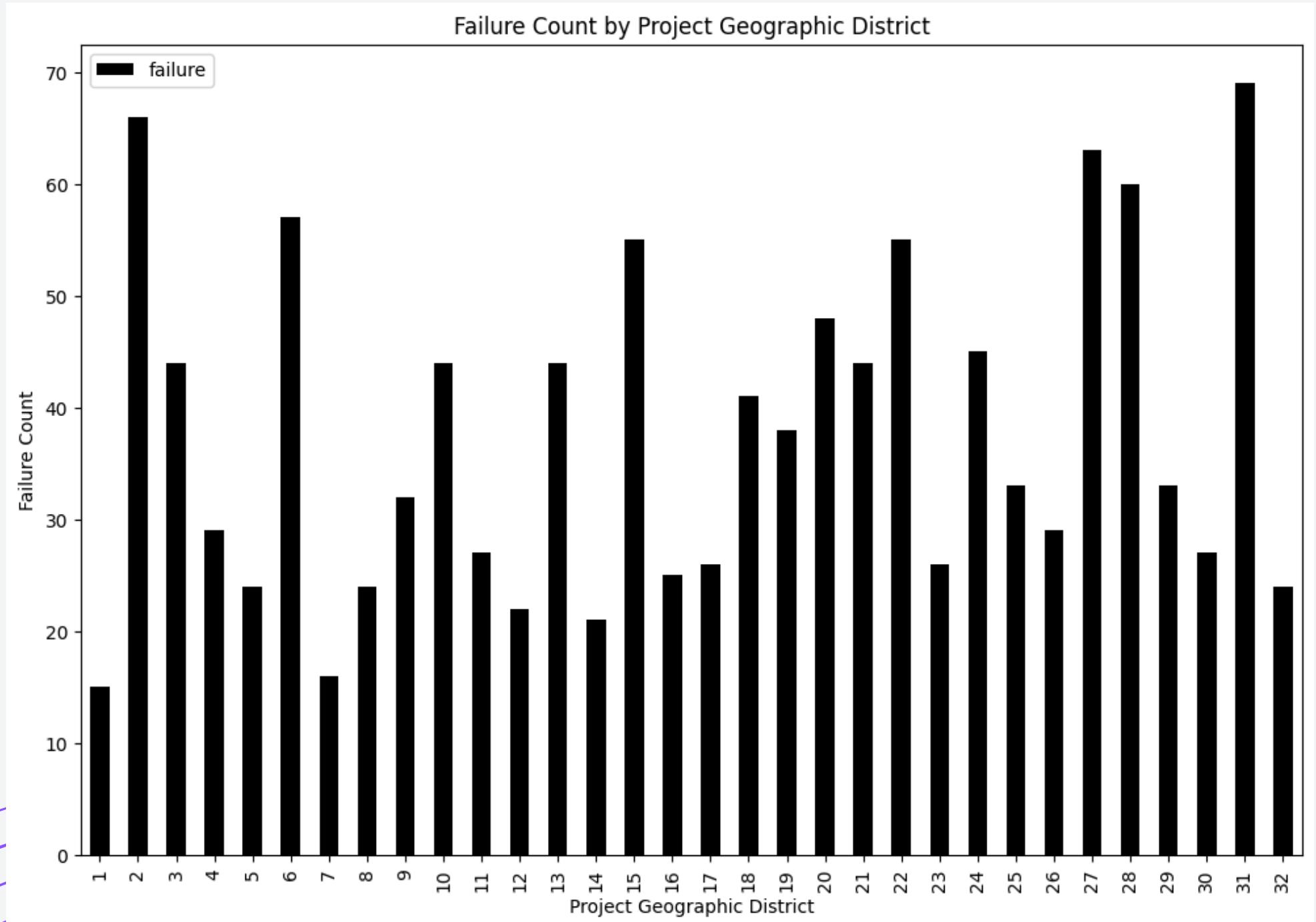
EXPLORATORY DATA ANALYSIS

CORRELATION OF FEATURES

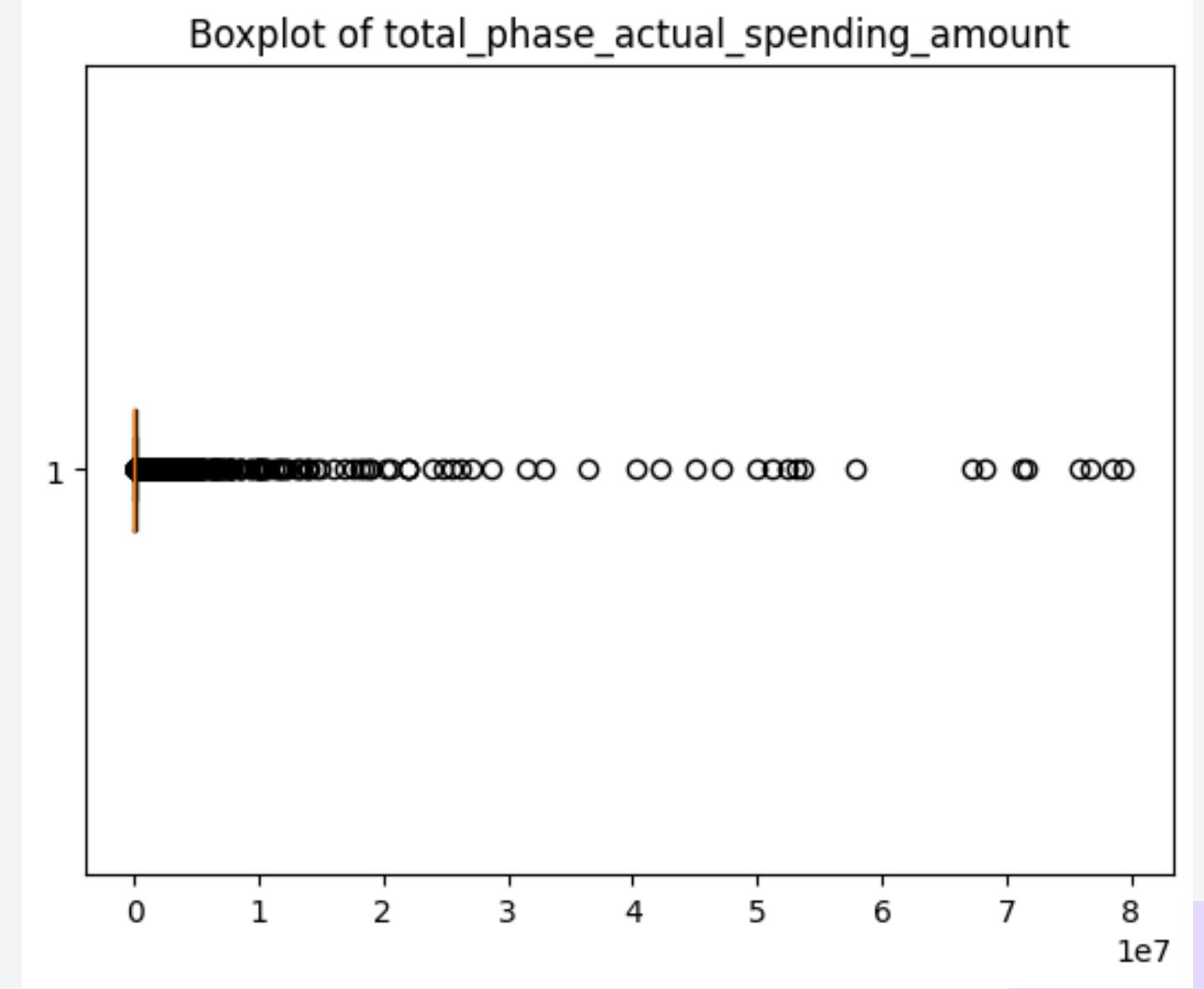
Feature	Correlation to failure
Location	-0.02
Duration	0.05
Excess budget	-0.1
Project type	0.18

Not enough significance in
correlation

EXAMINING SOME INDEPENDENT FEATURES

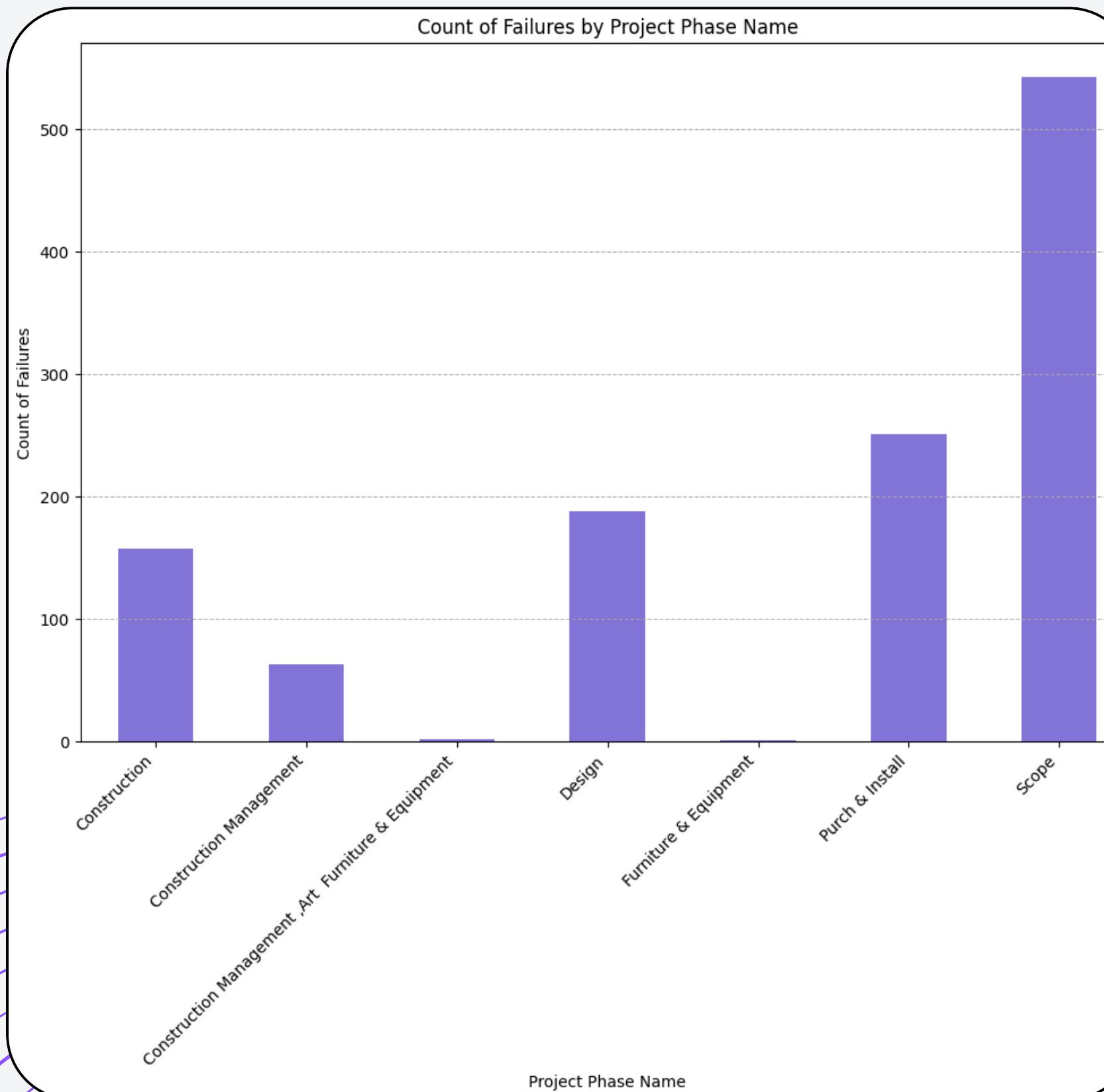


Some areas have more counts of failure than others



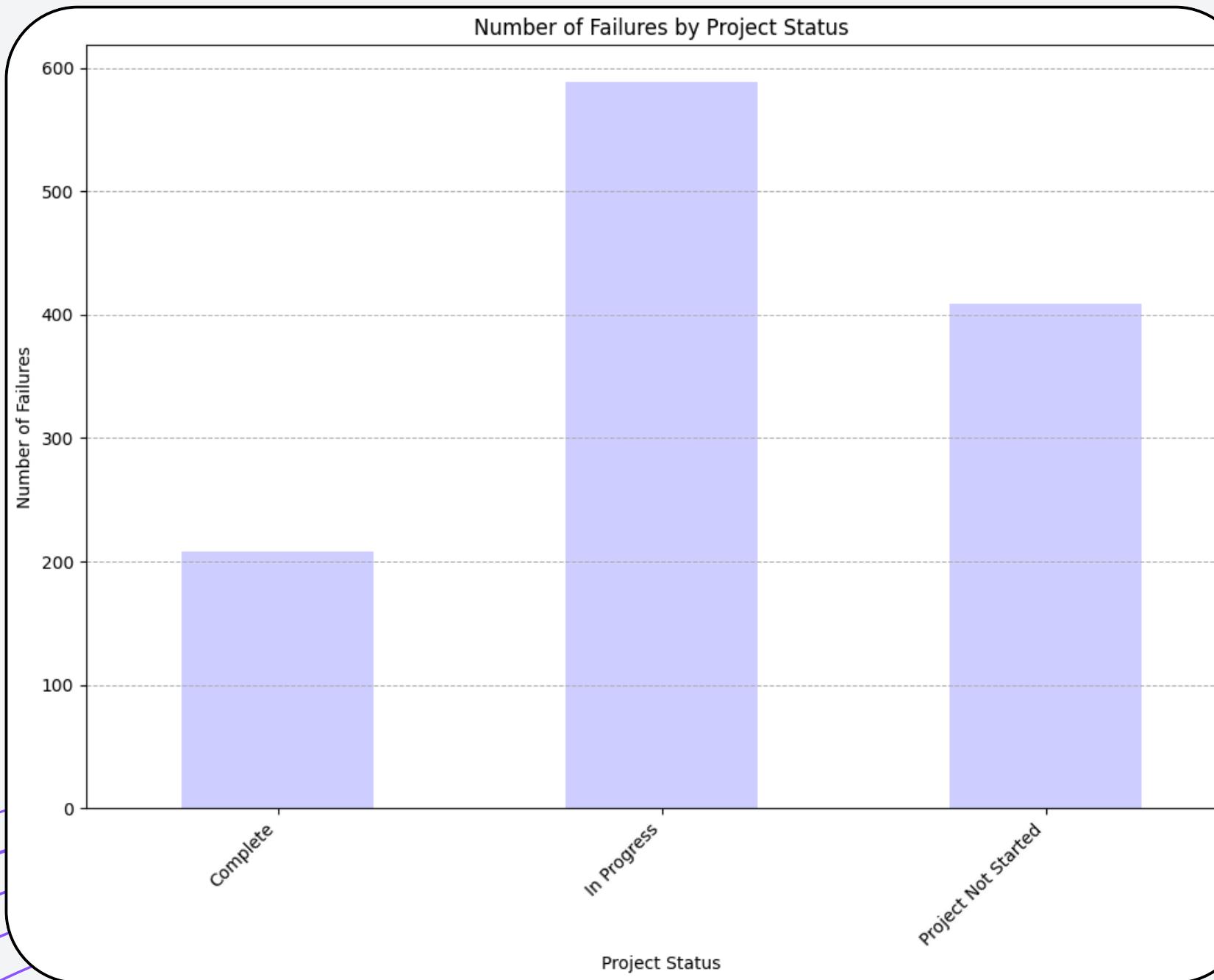
Very right skewed – spending a lot more than median

EXAMINING SOME INDEPENDENT FEATURES



Likely to have severely
underestimated initial costs – most
projects failed in scoping phase

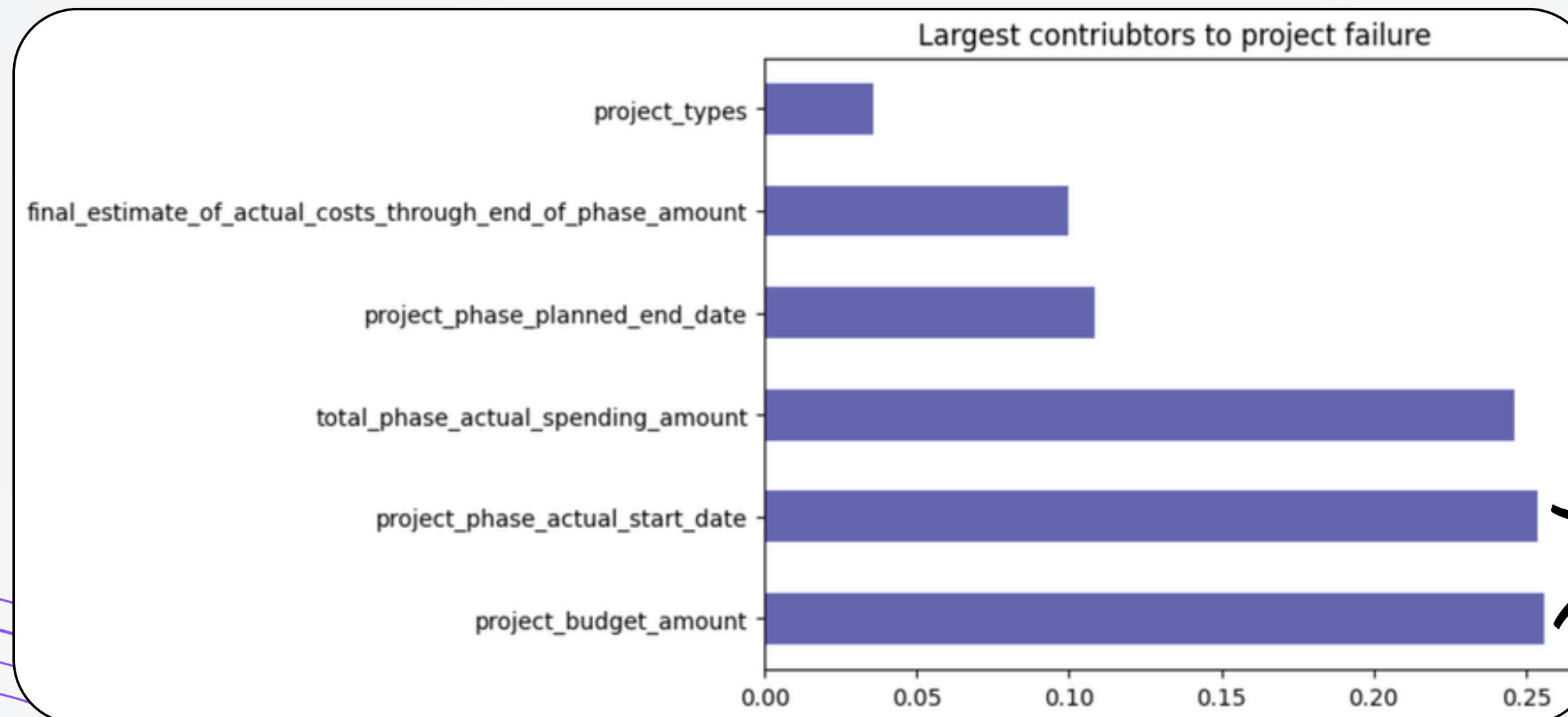
EXAMINING SOME INDEPENDENT FEATURES



Not started: Follows from project failure during scoping

In progress: Issues during development- delays, cost overruns, safety incidents

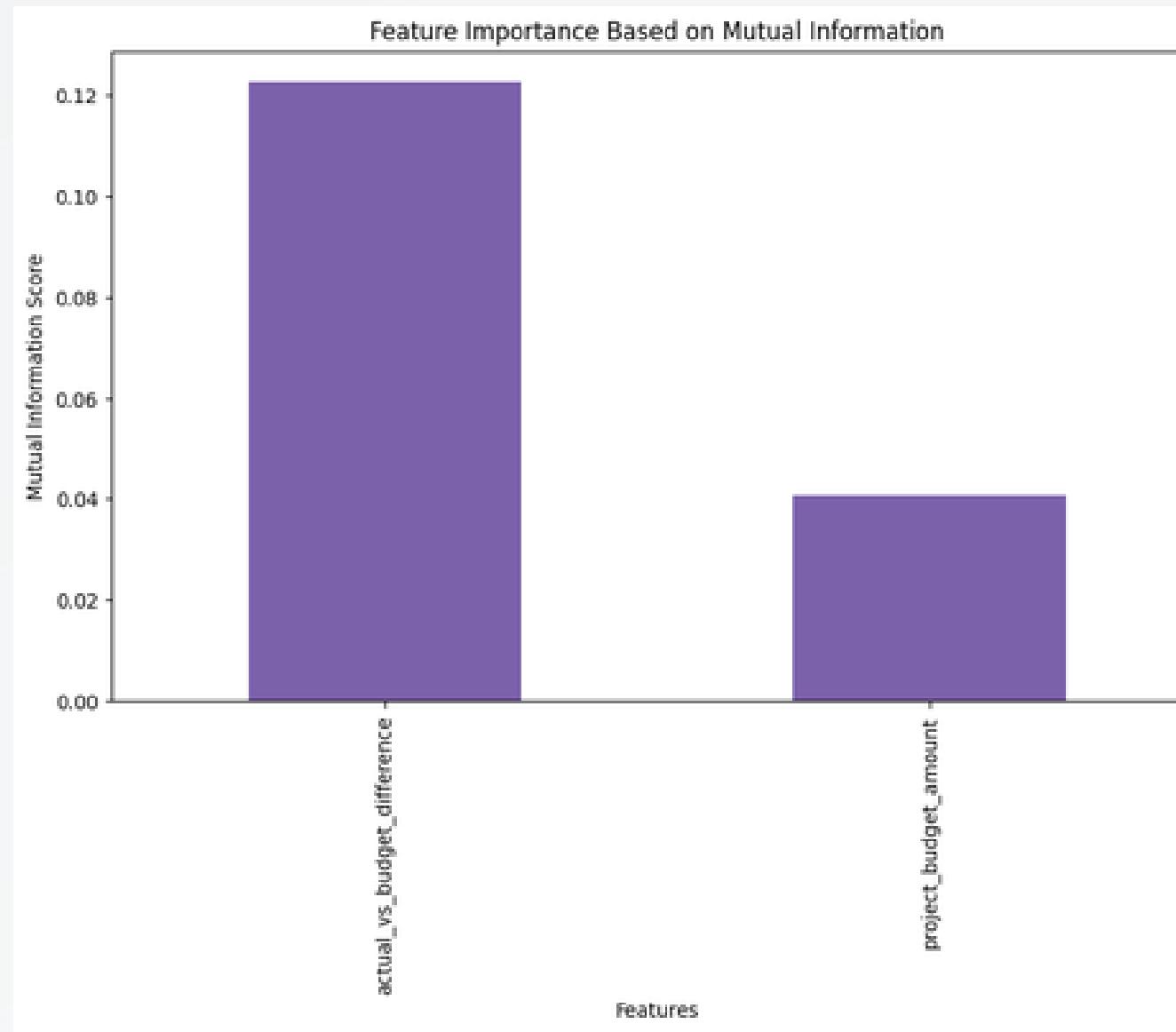
FEATURE IMPORTANCE



More recent projects have been failing more – economic factors

Higher budget than expected (likely during scoping) leads to failure in scoping phase

DEEPER LOOK INTO MOST IMPORTANT FEATURE

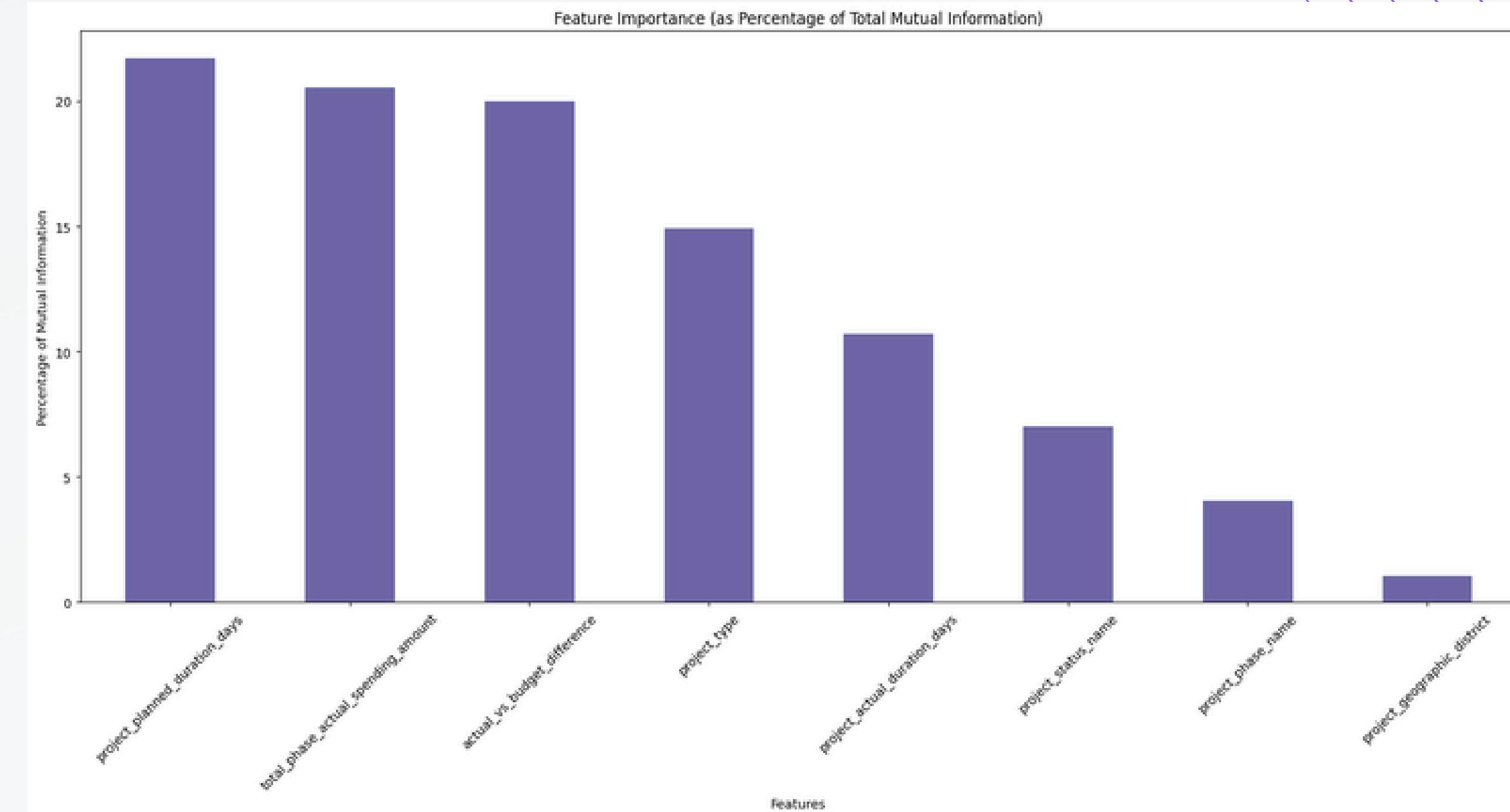


Excess amount spent was more of an issue than the budget itself – not very good investment

Removing the “actual budget” column then, duration turned out to be more important – the more time a project takes, the more likely it was to fail.

Concern: 20% importance is too low

Reason: Vastly different data scales, features not having strong relationships with the target variable



MODELLING LIKELIHOOD

**LOGISTIC
REGRESSION**
**91.4% MODEL
ACCURACY**

Feature	Coefficient
Project Budget	- 2.06e-06
Actual spending amount for current phase	8.68e-07
Estimate of costs for current phase	6.29e-07
Planned end date of current phase	- 1.09e-07
Actual start date of current phase	6.33e-08
Actual end date of current phase	4.46e-08
Project type	8.71e-11
Geographic district of project	9.49e-11
Intercept	2.02E-12

MODELLING LIKELIHOOD

RANDOM
FOREST

91% PRECISION
88% RECALL

Dealing with **imbalanced classification** – more non-failures than failures

Strategy: SMOTE (Synthetic Minority Over-sampling Technique) – decreasing the number of instances in the majority class by randomly removing instances.

RECOMMENDATIONS

- Only start scoping for project types with low project failure - more careful investments
- Be mindful of duration - the more delays for a project, the more likely it is to fail
- Be mindful of budgeting - avoid going over budget. Implement cost management strategies.
- use the model for future failure prediction

RAG

RETRIEVAL AUGMENTED GENERATION - EXTERNAL DATA RETRIEVAL
USING LARGE LANGUAGE MODELS AND GENERATIVE AI

Usage in context: Assist the client to efficiently search local building policies (zoning policies, environmental, etc.) while planning developments.

Model details and implementation in [Appendix](#)

EXAMPLE QUESTION

User

What is Green Mark Certification in
singapore_building_control_env_sus_regulations_
dc.pdf?

RAG agent

Green Mark Certification is a sustainability rating system and certification program in Singapore, primarily used for buildings. It is designed to promote sustainable building practices and improve the environmental performance of buildings. The certification is voluntary, but many buildings in Singapore aim to achieve Green Mark certification to demonstrate their commitment to sustainability.

EXAMPLE QUESTION

User

What does natural ventilation mean in Building Control Regulations 2003 Singapore_dc.pdf?

RAG agent

Natural ventilation means the supply of outside air to a building or removal of inside air from a building by means of windows and other openings.

APPENDIX

MODEL DETAILS

Model: ibm/granite-13b-chat-v2

Database: Chromadb

Embedding: WatsonxEmbeddings

PDF Loader: PyPDFLoader

Splitter: CharacterTextSplitter

CODE IMPLEMENTATION

```
# from singapore_building_control_env_sus_regulations_dc.pdf
question = "what is Green Mark Certification in singapore_building_control_env_sus_regulations_dc.pdf?"
response = rag_query(question, retriever, model)
# from Building Control Regulations 2003 Singapore_dc.pdf
question2 = "what does natural ventilation mean in Building Control Regulations 2003 Singapore_dc.pdf?"
response2 = rag_query(question2, retriever, model)
print(response)
print()
print(response2)
```

✓ 11.2s

Green Mark Certification is a sustainability rating system and certification program in Singapore, primarily used for

Natural ventilation means the supply of outside air to a building or removal of inside air from a building by means of

INTERFACE

Chatbot 

What is Green Mark Certification in singapore_building_control_env_sus_regulations_dc.pdf?

Green Mark Certification is a sustainability rating system and certification program in Singapore, primarily used for buildings. It is designed to promote sustainable building practices and improve the environmental performance of buildings. The certification is voluntary, but many buildings in Singapore aim to achieve Green Mark certification to demonstrate their commitment to sustainability.

What does natural ventilation mean in Building Control Regulations 2003 Singapore_dc.pdf?

Natural ventilation means the supply of outside air to a building or removal of inside air from a building by means of windows and other openings.

Very good answer!

  Type your message here