AUCA SIDP

STUDENT INITIATIVE DEVELOPMENT PROGRAM

powered by

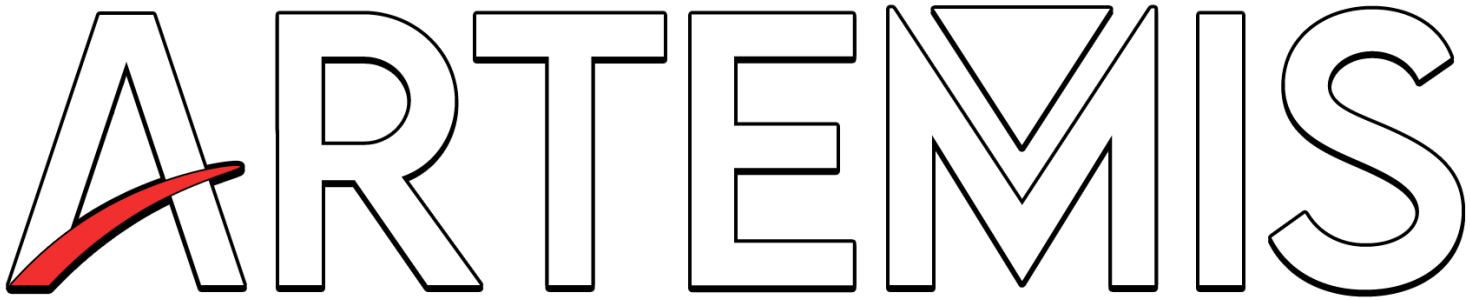
ARTEMIS

real-time management system

developed by **Luna Maltseva** *for* **SIDP**
in collaboration with **AUCA CCE**;
under supervision and guidance of
Nurzhamal Karamoldoeva, Aliia Iusupova;
as well as **Jonathan Becker** *and* **Erin Cannan**

Center for Civic Engagement

@aucq_cce



Grant Program Result Prediction

a Real-Time Management System to Estimate Results for a Given Budget

Overview → Statement → Model → Data → Scripts → Result

Agenda For The Presentation

Phase I: Application

Phase II: Realization

Student

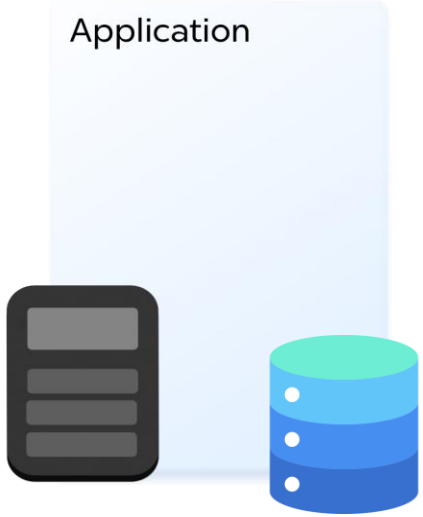
Project leaders submit project documentation, which includes descriptions, strategy, estimations, and budget

Upon completion, project leaders provide a detailed report of the project's results in addition to substantiating documents

Staff

After the call has been closed, the SIDP Committee executes the selection process in tandem, greenlighting projects and allocating budgets

Phase I: Application



Phase II: Realization

Student

Project leaders submit project documentation, which includes descriptions, strategy, estimations, and budget

During the project, leaders regularly report key metrics

Upon completion, project leaders provide a detailed report of the project's results in addition to substantiating documents

Staff

After the call has been closed, the SIDP Committee executes the selection process in tandem, greenlighting projects and allocating budgets

As projects are being realized, the SIDP Committee monitors the progression of each project, supporting project leaders and stepping in when necessary

Problem Statement

Mathematical Models

TF-IDF

$$D = \{d_1, d_2, \ldots, d_N\}$$

$$V = \{t_1, t_2, \ldots, t_M\}$$

$$tf_{ij} = \frac{f_{ij}}{\sum_{k=1}^{M} f_{ik}}$$

$$idf_j = \log \frac{N}{1 + n_j}$$

$$X_{ij} = tf_{ij} \times idf_j$$

Ridge Regression

$$X = \{1, x_1, x_2, \ldots, x_p\},\ X \in \mathbb{R}^{n \times p}$$

$$\omega = \{\omega_0, \omega_1, \omega_2, \ldots, \omega_p\},\ \omega \in \mathbb{R}^{n \times p}$$

$$y \sim \hat{y} = X\omega + \varepsilon,\ y \in \mathbb{R}^n$$

$$\varepsilon \sim N(0,\ \sigma^2 I)$$

$$\hat{\omega}_\lambda = \arg\min_{\omega \in \mathbb{R}^p} \|X\omega - y\|_2^2 + \lambda \|\omega\|_2^2$$

Data Mining

$$p(y|x) = \frac{1}{Z(x)} e^{\theta^\top F(x,y)}$$

$$Z(x) = \sum_{y'} e^{\theta^\top F(x,y')}$$

$$\hat{B} = \arg\max_B \sum_{(n,r) \in B} -|pos(n) - pos(r)|$$

$$p(y|s_i) = \frac{1}{1 + e^{-\omega^\top \varphi(s_i)}}$$

$$\hat{\theta} \pm z_{a/2} \sqrt{V(\hat{\theta})}$$

Finite Multidimensional Dynamic Programming

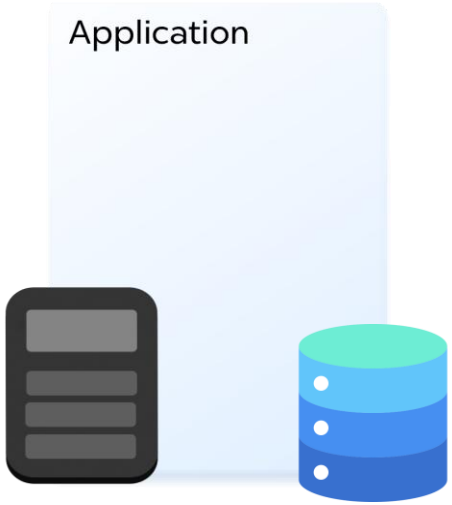
$$s' = f(s, a, \xi)$$

$$V_t(s) = \min_{a \in A(s)} \mathbb{E}\left[c_t(s, a) + V_{t+1}(s')\right]$$

Obtaining Cleaned Data for Regression/Numerical Tests

[39 rows x 68 columns]

ARTeMiS

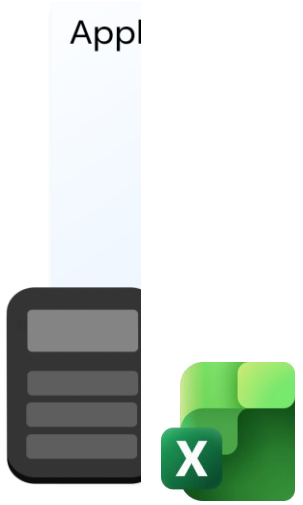


Raw

Clean

[39 rows x 5 columns]

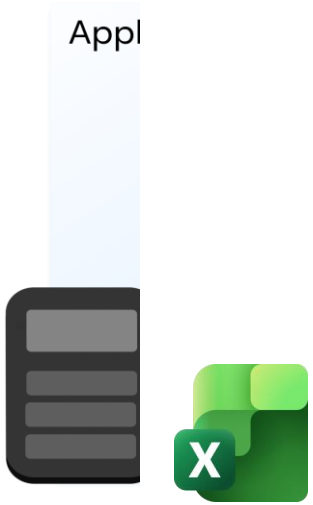
Regression*



Cleaned

[39 rows x 7 columns]

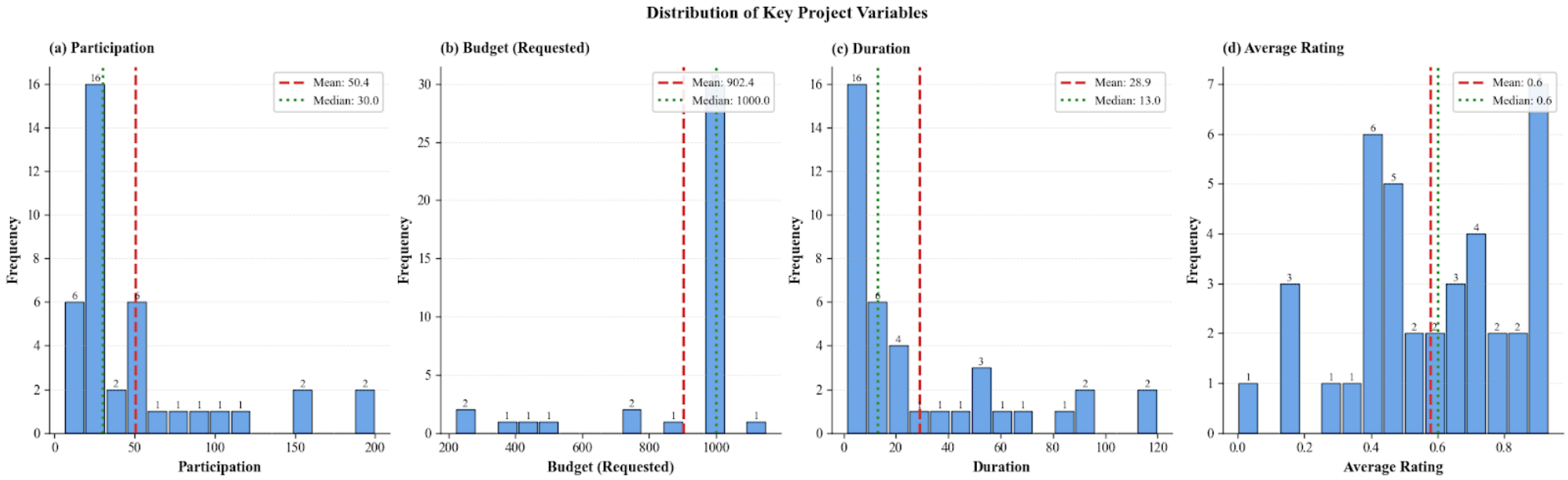
Numeric

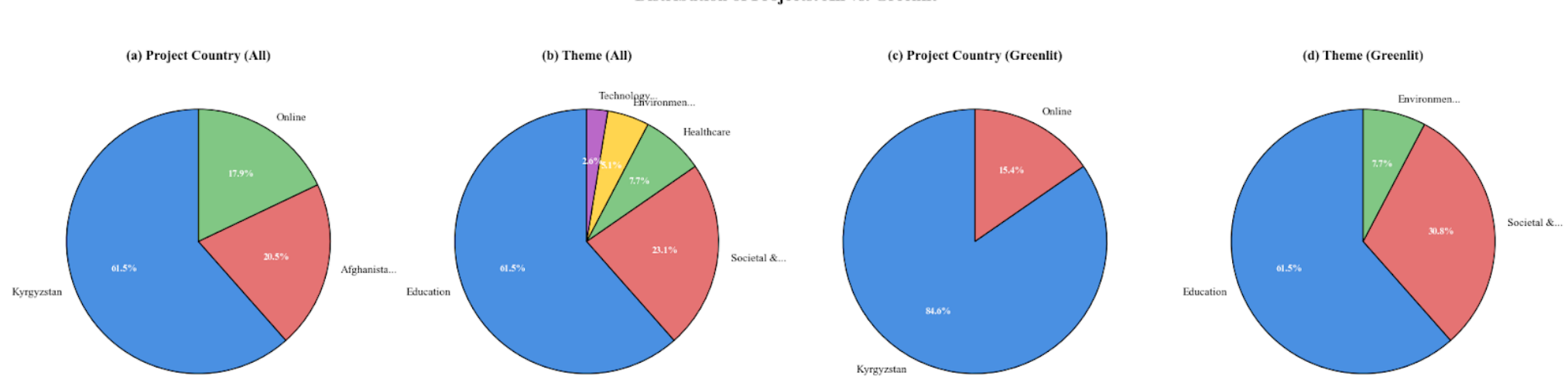


Cleaned

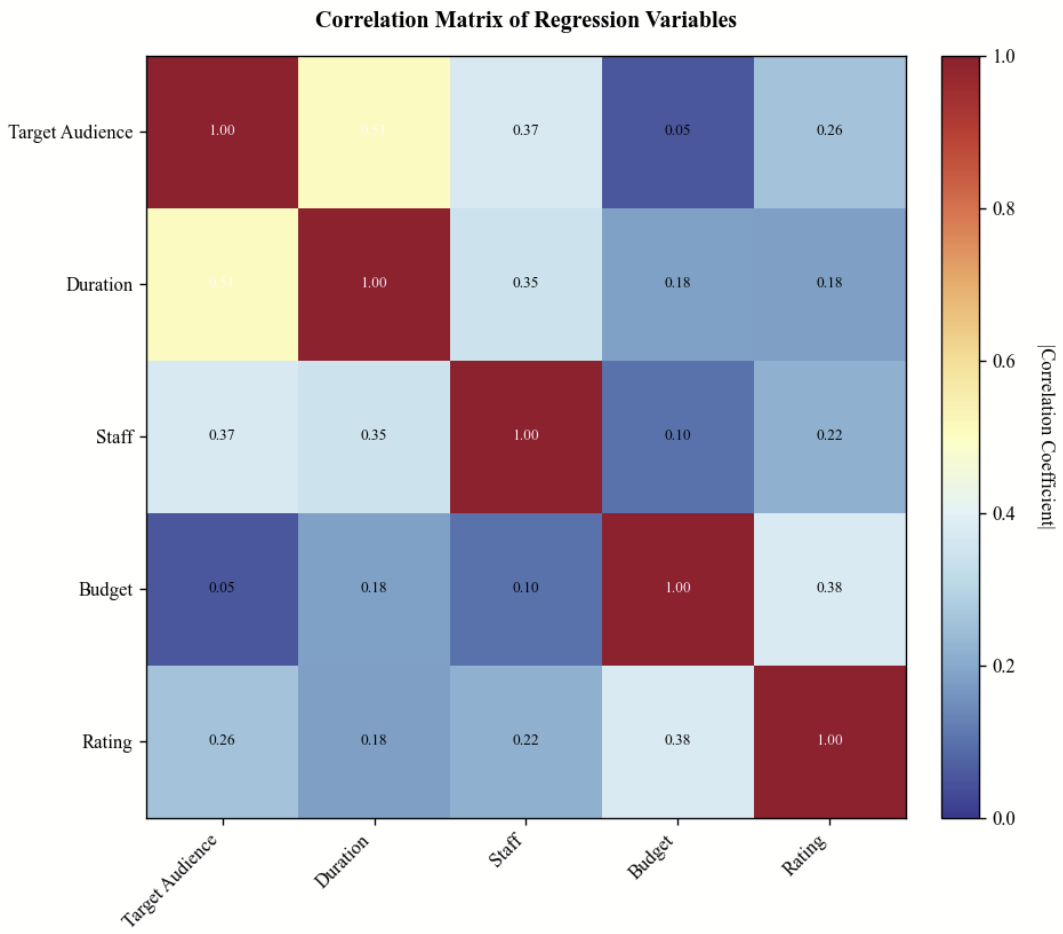
*only numeric data has been used for the regression model in order to shrink the total amount of features down to the possible minimum, given that the dataset only has 39 entries

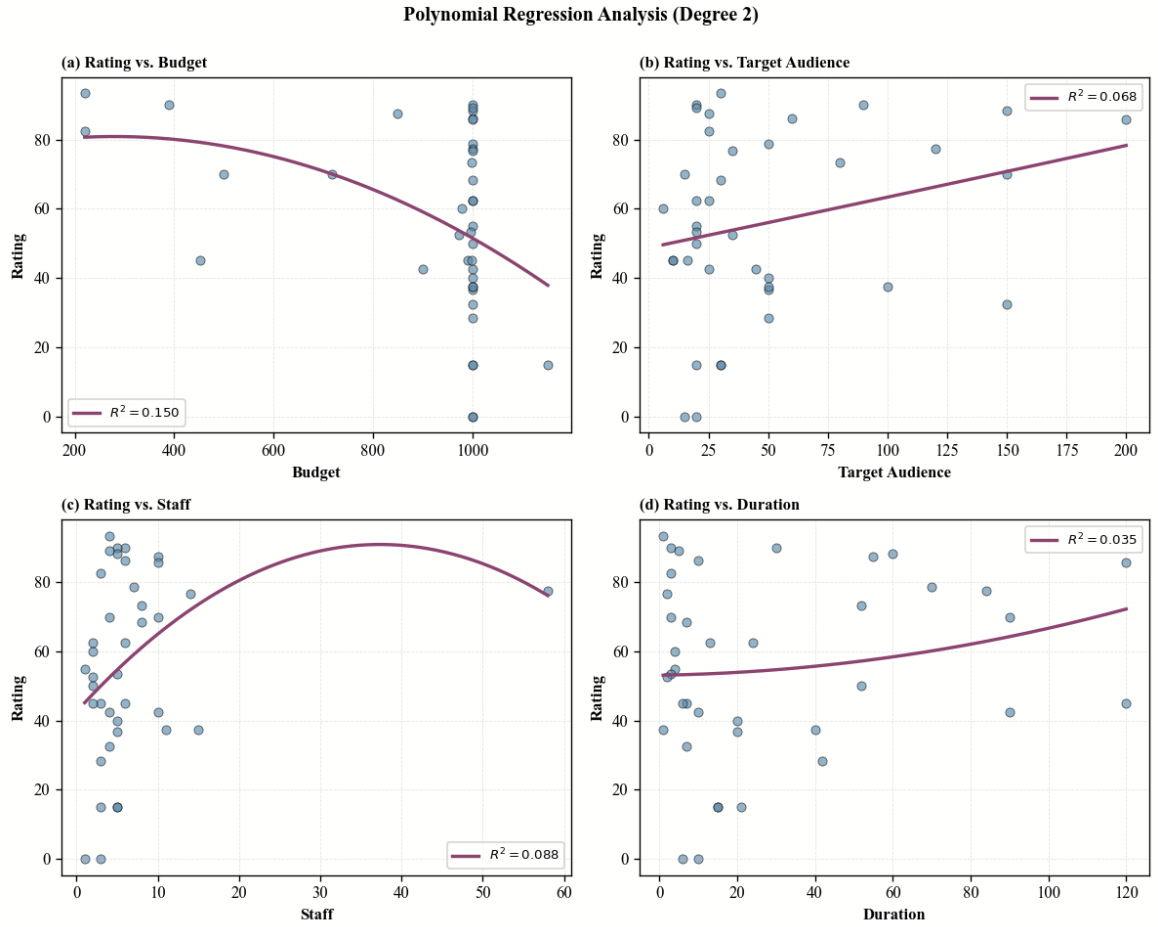
Numeric Data Visualized





Regression Data Visualized

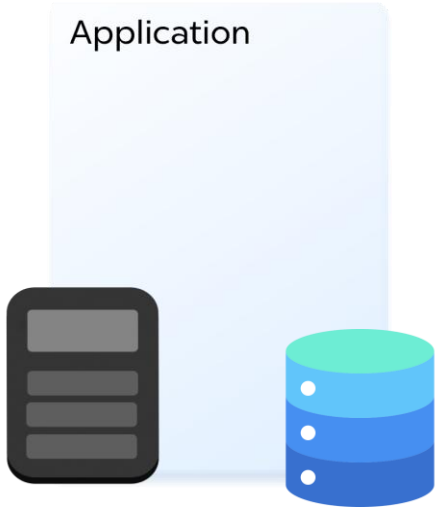




Obtaining Cleaned Data for Dynamic Programming

[39 rows x 68 columns]

ARTeMiS



Raw

Data Mining

21.8%*

*barely usable data

[60] [19]

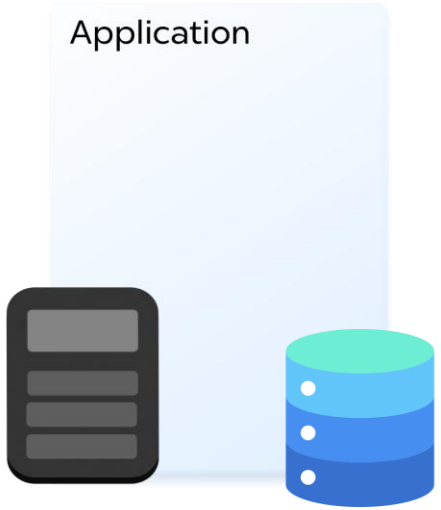
2024 Fall & 2025 Spring

Raw

Obtaining Cleaned Data for Dynamic Programming

[39 rows x 68 columns]

ARTeMiS



Raw

Confidence Interval

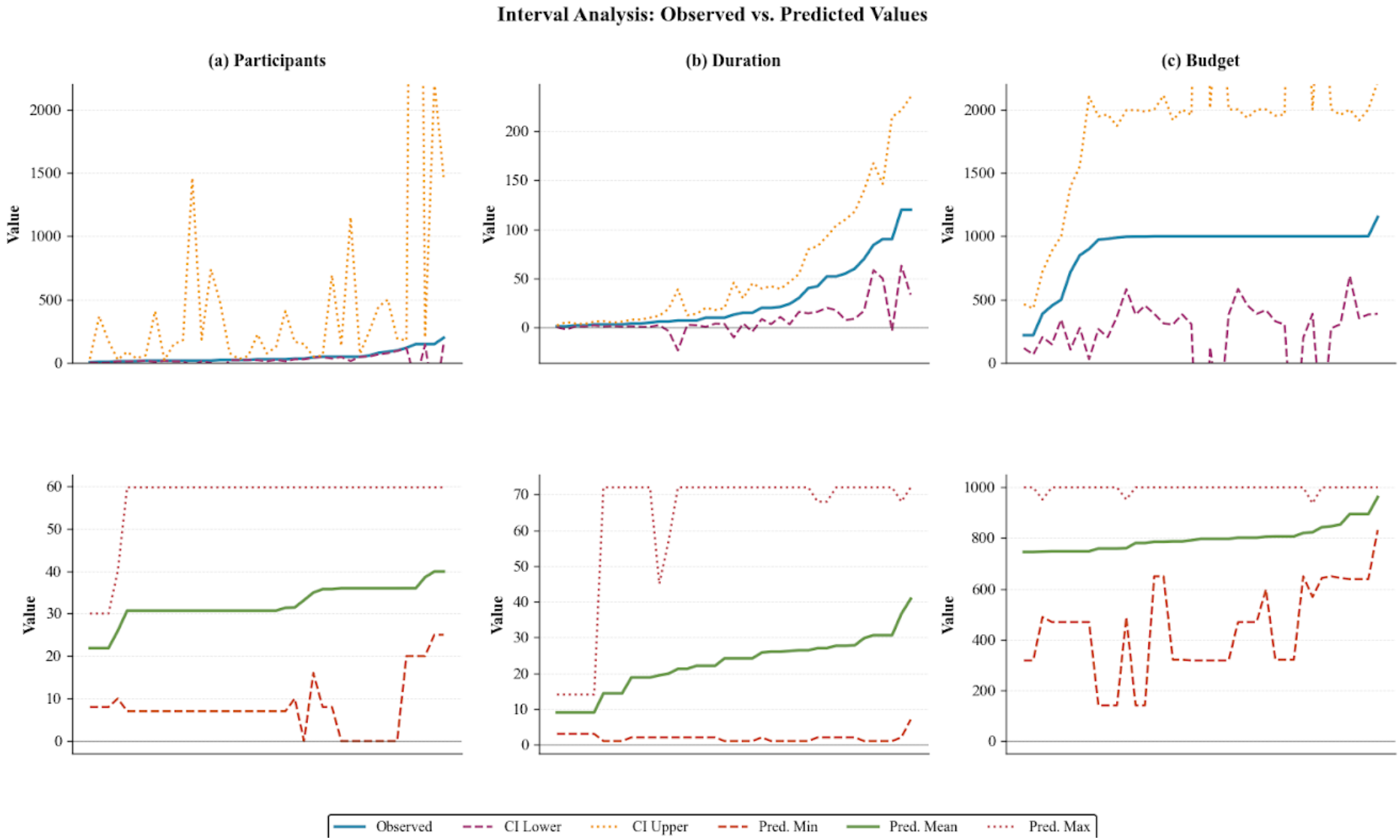
*only implemented projects

[106 rows x 20 columns]

SIDP/ACEP

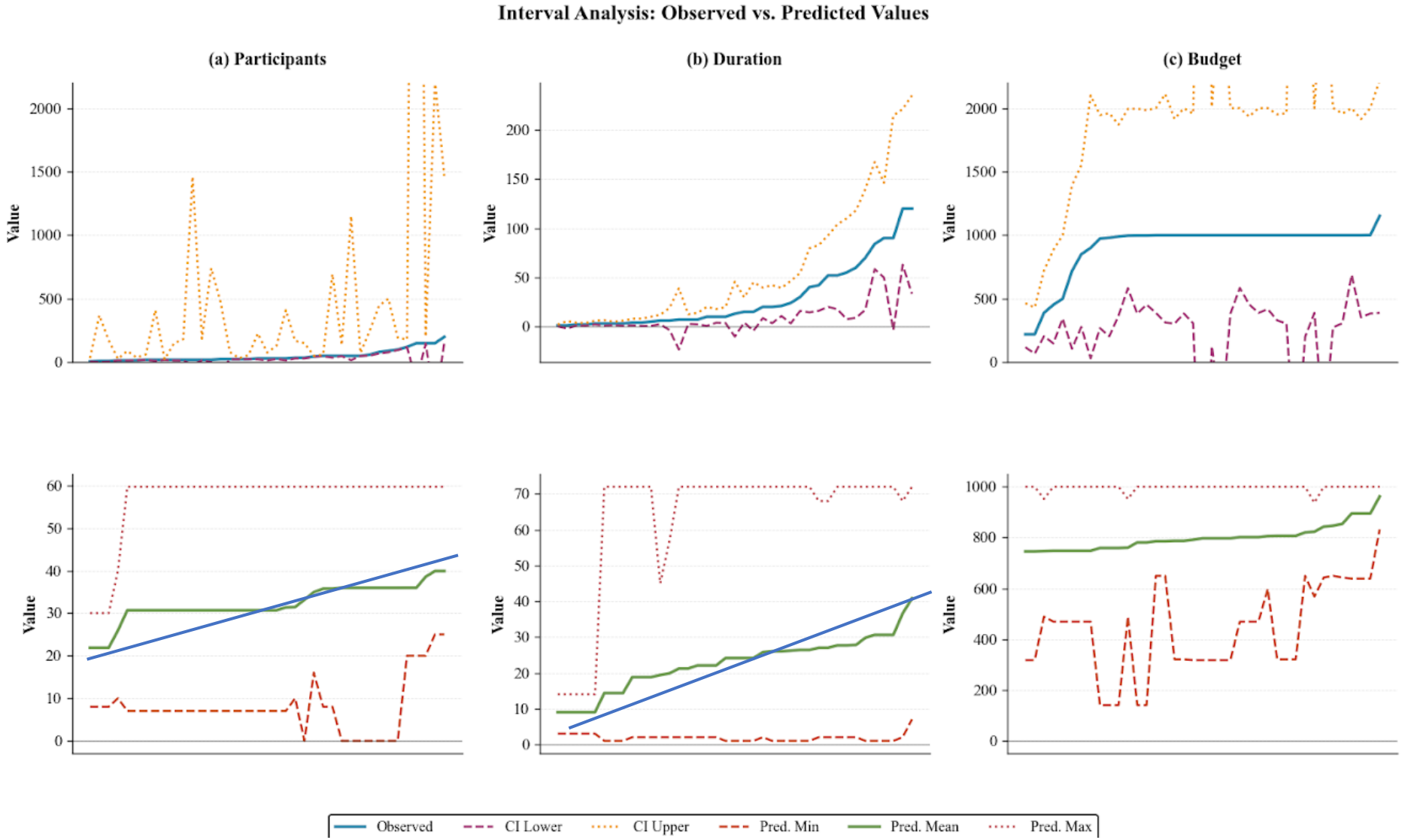
Cleaned*

Interval Data Visualized



Interval Data Visualized

Normalized w/ Rating



Regression

```
15 model = Pipeline([
16     ("poly", PolynomialFeatures(degree=2, include_bias=False)),
17     ("scaler", StandardScaler()),
18     ("ridge", Ridge(alpha=10))
19 ])
...
23 model.fit(X_train, y_train)
24 test_r2 = r2_score(y_test,model.predict(X_test))
```

```
CV R2   = -0.145
Test R2 = -0.083
```

Finite Multidimensional Dynamic Programming

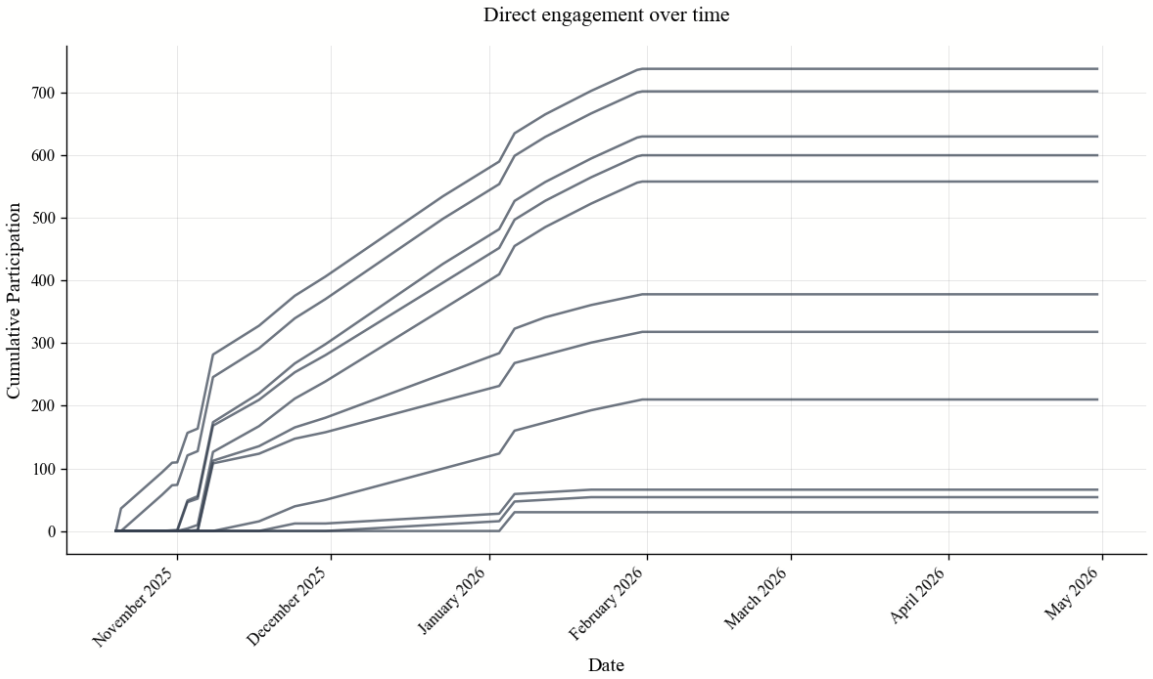
```
9  def select_projects_dp(
10       filepath: str,
11       max_budget: int,
12       theme_diversity_factor: float,
13       country_diversity_factor: float,
14       max_states: int = 200_000,
15       verbose: bool = False ) -> List[int]
...
37       for idx, it in enumerate(items):
...
39               for (sel_count, budget_used, counts_tuple), (obj_tuple, sel_ids) in snapshot:
...
85       for (sel_count, budget_used, counts_tuple), (obj_tuple, sel_ids) in dp.items():
...
193 incoming = [1, 2, 3, 4, 11, 13, 22, 24, 29, 32, 36]
...
195 selected = select_projects_dp(
         path = "../data/artemis/artemis_data_for_DP.xlsx",
         max_budget = 9700,
         theme_diversity_factor = 0.8,
         country_diversity_factor = 0.95,
         max_states=200_000)
```

Results

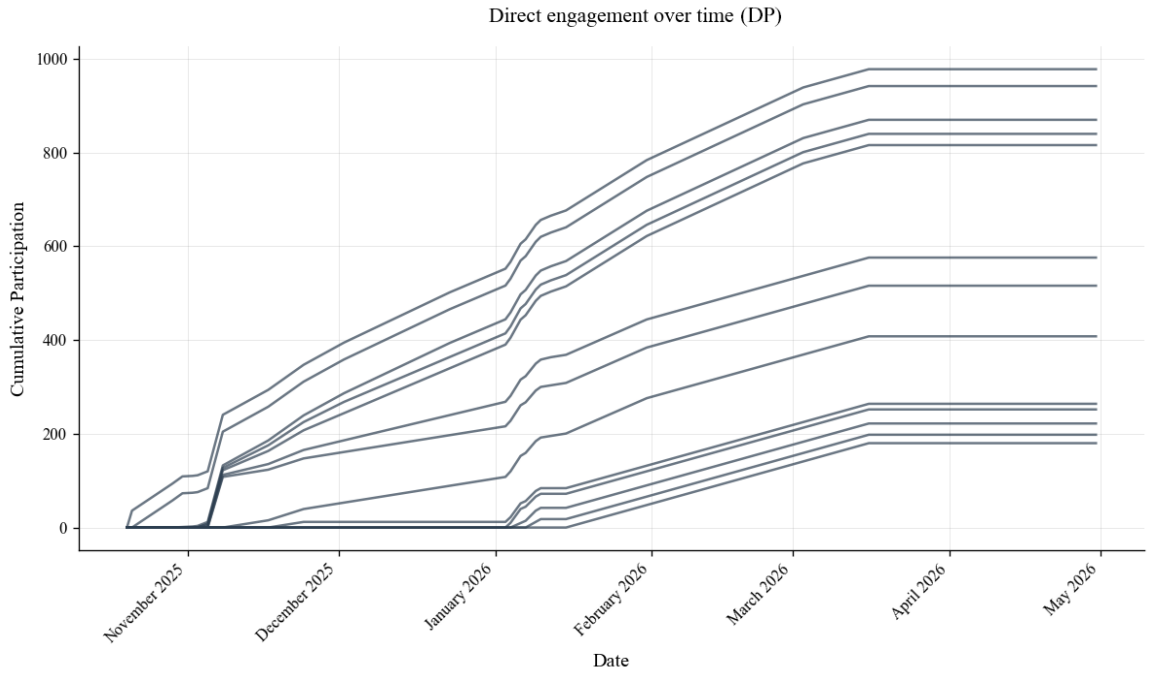
Overlap similarity: 72.7%

Immediate improvement: 129.3%

Selected Manually



Selected via Algorithm



Center for Civic Engagement

@auca_cce

THE FLOOR IS OPEN TO QUESTIONS

by Luna Maltseva

AUCA SIDP

STUDENT INITIATIVE DEVELOPMENT PROGRAM

powered by

ARTEMIS

real-time management system

developed by **Luna Maltseva** *for* **SIDP**
in collaboration with **AUCA CCE**;
under supervision and guidance of
Nurzhamal Karamoldoeva, Aliia Iusupova;
as well as **Jonathan Becker** *and* **Erin Cannan**