

RunBuggy I4X Presentation

Building a Scoring Model for Vehicle Repossession Sites Using Web-Scraping And RunBuggy Order Data



Noah Golder, Yuxing Liu, Vaidehi Sinha, Leonardo Bangayan, Richard Wang, Alyssa Le
The Basement at the University of California, San Diego

Project Background

- **RunBuggy** is an open marketplace and technology platform that connects car shippers and haulers to make shipping cars faster, easier, and smarter.
- By eliminating the hassle and reducing the cost of shipping a car, RunBuggy makes the experience of moving a car from point A to point B – better.

Problem Statement

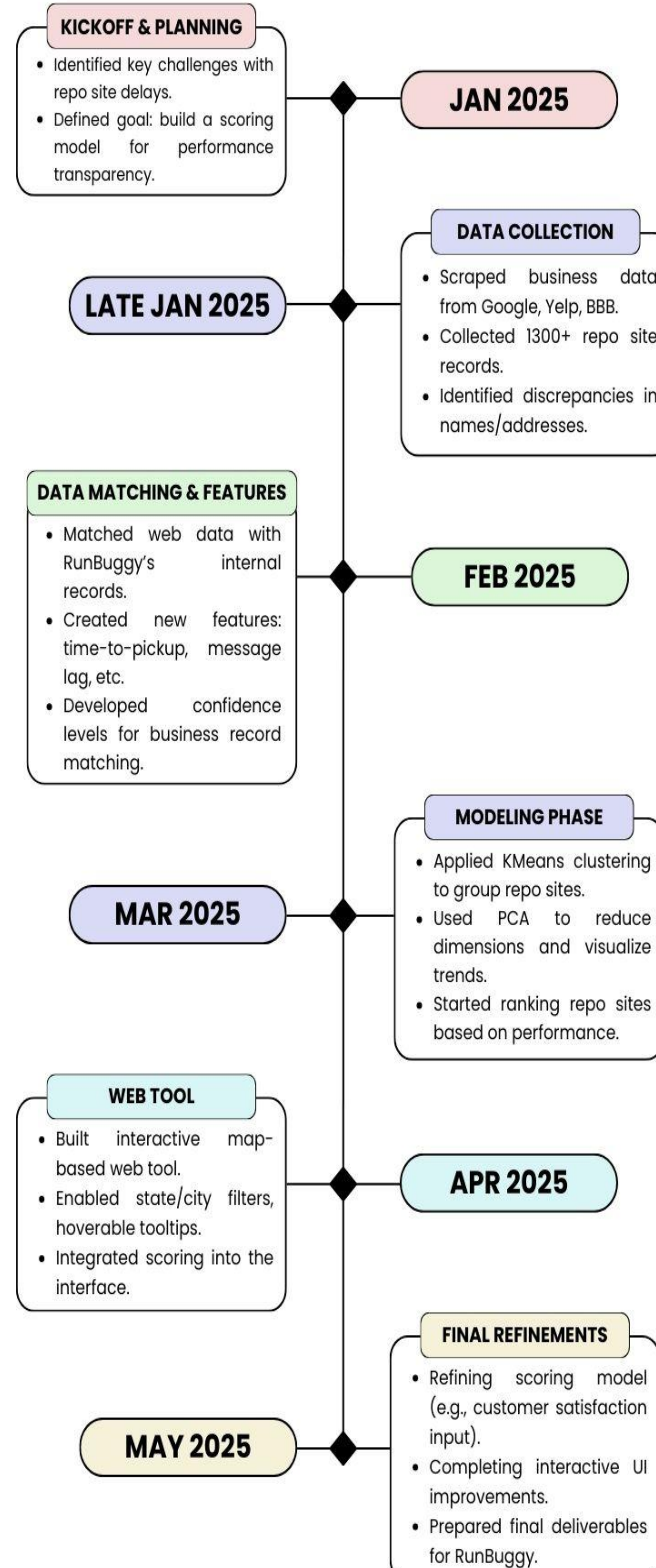
Repo sites, where repossessed vehicles are stored and sold, often create issues for RunBuggy, lenders, and towed individuals. We aim to use data to bring transparency and improve customer experiences.

Project Aims and Objectives

- **Improve Repo Site Transparency:** Repo sites often cause issues for RunBuggy, lenders, and customers. Our goal is to use data to bring clarity and visibility into repo site operations to improve overall experience.
- **Transform Unstructured Data:** Much of the customer feedback exists in text form. We aim to apply LLMs and Python to clean and convert this into usable, numerical data that can be analyzed and visualized.
- **Score Repo Site Quality:** Our objective is to build a model that classifies repo sites as “good” or “bad,” similar to Yelp ratings for restaurants. This will help users and businesses make more informed decisions based on data.



Project Journey and Milestones

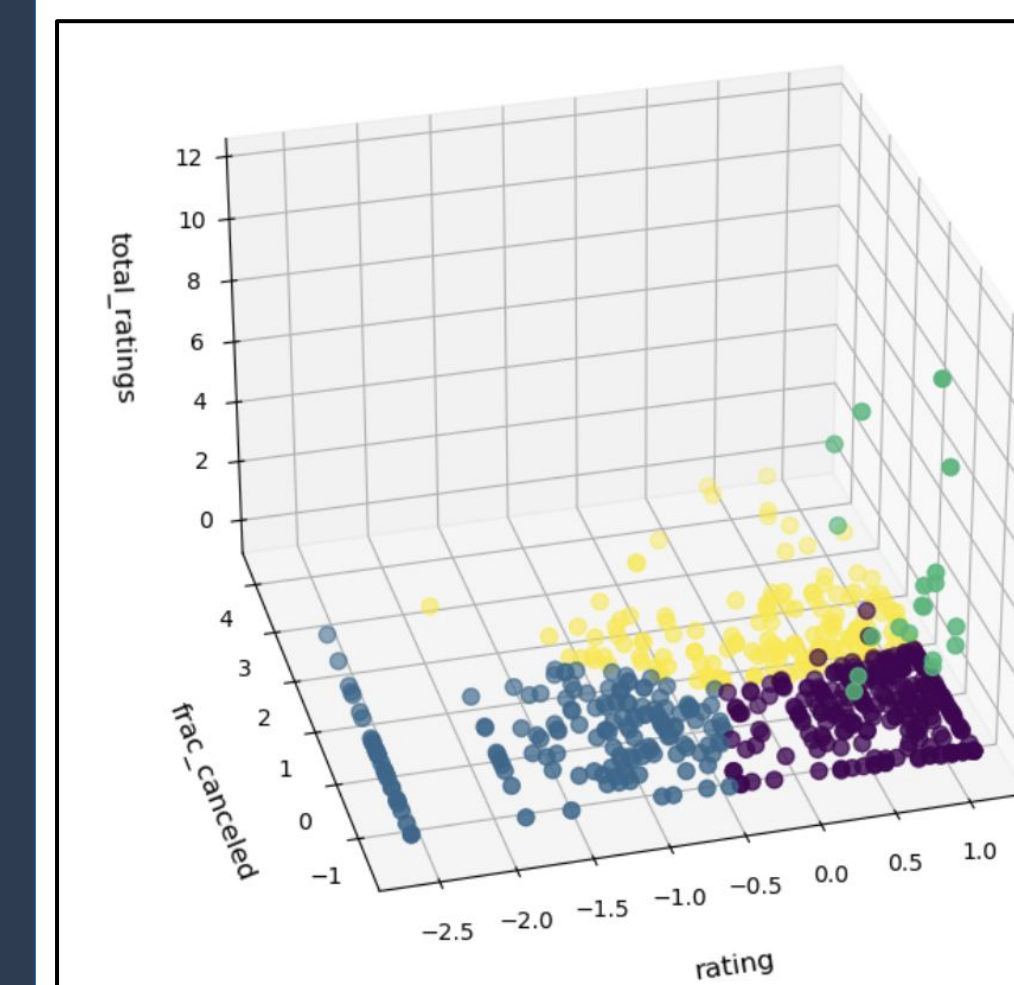


Solutions and Deliverables

Computing Repo Site Scores: For the 500 sites in our confidence-scored dataset, we used unsupervised learning models to produce discrete and continuous scores.

Clustering Model:

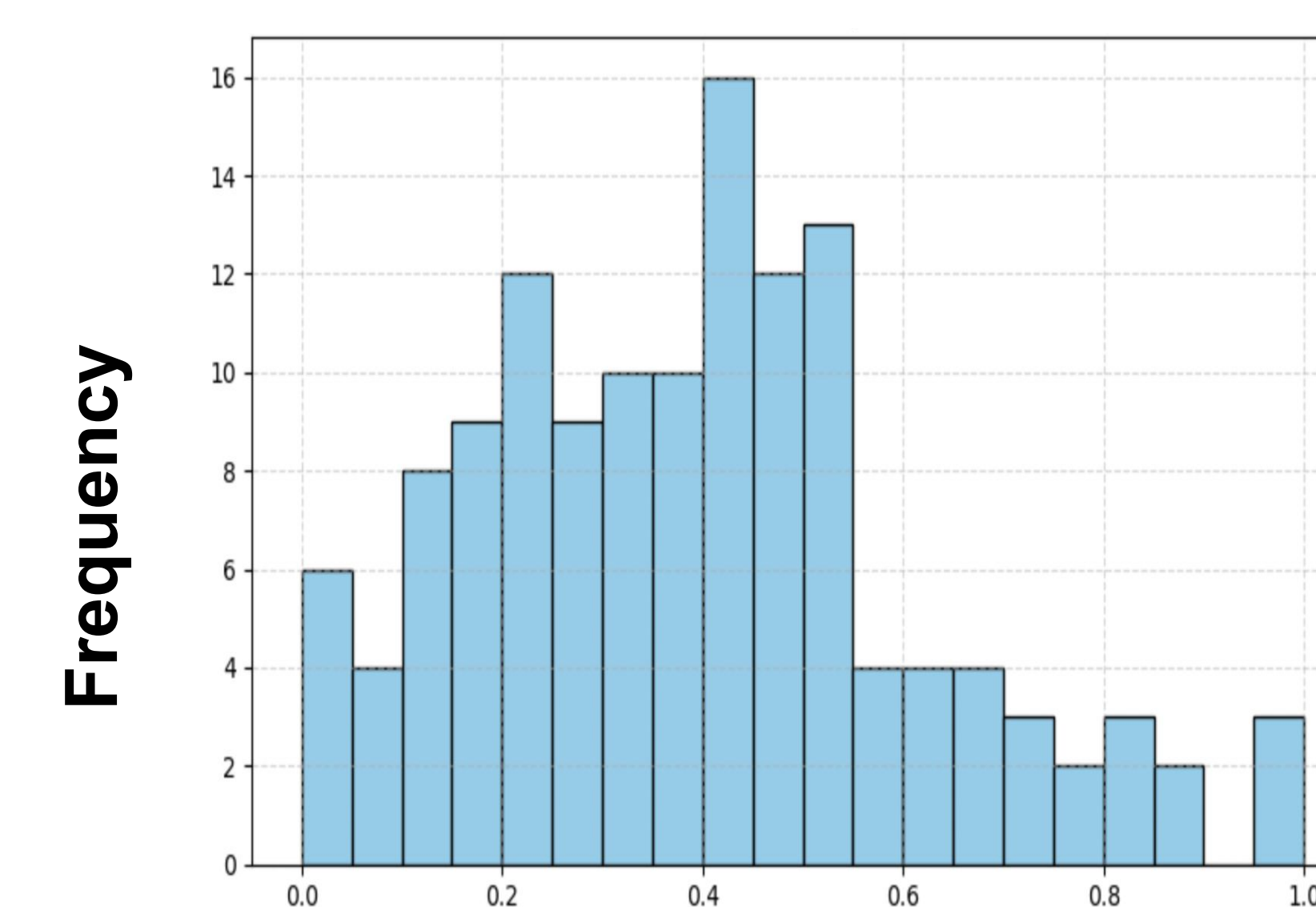
For the clustering model, we calculated discrete repo site scores by grouping the sites using k-means clustering. We came up with 4 categories of sites based on three features – total Google ratings, fraction of orders cancelled, and Google rating (0.0 – 5.0).



	Total ratings	Frac_cancelled	Rating
Awesome companies	+	+	+
Promising gems	-	+	+
Risky business	-	-	+
Just bad	-	-	-

Principal Component Analysis (PCA) Model:

Since the clustering model does not account for variations within clusters, we computed a continuous score using pca, which reduces our high-dimensional data to a singular score per repo-site. Below is the distribution of repo scores using features total_google_ratings, google_rating, total_yelp_ratings, yelp_rating, and proportion_customer_message.



PCA Score – Customer Satisfaction

Learning Outcomes

Data Integration and Insights: We cleaned and merged internal datasets with public data to produce actionable insights

User-Centered Thinking: We translated data-driven findings into clear, accessible insights aligned with business priorities.

Applied Statistical and Machine Learning Techniques: We engineered features, applied clustering algorithms and scoring models to identify patterns across locations.

Real-World Problem Solving: We developed and refined evaluation methods while navigating real-world ambiguity, iterating on models and adapting to evolving project goals.

Essentials of Entrepreneurship: We gained exposure to core business concepts such as intellectual property, business modeling, and design thinking.

Next Steps for RunBuggy

Looking ahead, the project can be strengthened by implementing real-time data updates through automated pipelines that regularly pull updated business information from platforms like Google and Yelp. Enhancing the filtering features would allow users to refine search results based on criteria such as rating, confidence level, and geographic proximity. Improving the map-based user interface with interactive elements—such as clustering, hover details, and match highlights—would further enhance usability. Introducing a user feedback mechanism where visitors can suggest corrections or validate business matches would support continuous data quality improvement. Additionally, developing internal dashboards for RunBuggy teams could offer valuable insights.

Video QR Code and Contact Info

Learn more about this I4X project



Team Information:

Noah Golder
(ngolder@ucsd.edu)
Yuxing Liu (yul269@ucsd.edu)
Vaidehi Sinha
(vsinha@ucsd.edu)
Leonardo Bangayan
(lbangayan@ucsd.edu)
Richard Wang
(riw009@ucsd.edu)
Alyssa Le (a6le@ucsd.edu)