Home agAIn
Capstone Final Presentation

Michelle Sinani, Siddharth Chandras, Mu'izz Widatalla, Jonathan Mulieri, Scott Thompson
10 December 2024

Image source: link.

Team Home AgAIn

Michelle Sinani
Data Analyst and Subject Matter Expert

Jonathan Mulieri
Software Architect and Entrepreneur

Siddharth Chandras
Engineer and API Developer

Mu'izz Widatalla
Data Engineer and Consultant

Scott Thompson
Designer and Data Analyst

Problem → Solution → Impact

Problem

Many people exit homelessness programs without permanent housing.

Scale of problem: Over half a million people are homeless in the US[1]

Solution

Learn patterns of successful services. Recommended individualized service plans.

Key user persona: Case Managers, who coordinate services for homeless individuals[2]

Impact

Improve lives and help restore confidence in homelessness organizations.

Huge potential impact: Over 10,000 homelessness organizations[3] in US

1. The U.S. Department of Housing and Urban Development (HUD). Link.
2. Weightman, et. al (2023) citing Ponka (2020). Link.
3. IBIS World. Link.

Image Sources: Link, Link, Link

Where we are going...

Hypothetical example of a participant

Alex Johnson

Participant Information

ID: CASE-1

Age: 34

Gender: Transgender

Race: White

Prior Living Situation: Emergency Shelter

Disability Status: Physical Disability

Chronically Homeless: Yes

Has Dependent(s): No

Has Pet(s): Yes

Is Veteran: Yes

Completed High School: No

Entry Income: $500

Program Entry Date: 2024-11-05 (29 days)

Services Provided

Add Service

Service Name	Count	Actions
Basic Needs Service	1	
Disability Support Service	2	

Services provided to participant

Where we are going…

Customized Service Recommendations

Service Recommendations

Permanent Housing Likelihood Improvement: **52%** (0.44196 → 0.95747)

Pet Support Service
Services supporting participants with pets, ensuring the well-being of both the individual and their animal companion.

Veteran Services Service
Specialized support addressing the unique needs of veterans, including benefits and reintegration services.

Education Support Service
Support for accessing educational resources and supplies, as well as guidance on furthering education and career prospects.

Housing Search Service
Guidance in searching for and identifying suitable housing options for participants in need of stable accommodations.

Benefits Assistance Service
Assistance with navigating, applying for, and securing benefits to help participants access necessary resources for stability.

Financial Support Service
Help covering fees, arrears, deposits, and other financial needs to provide stability and address participants' urgent monetary concerns.

IMPORTANT DISCLAIMER *The following services were custom generated for the participant and were identified as increasing their likelihood for a permanent housing outcome. These recommendations aim to support the case manager by efficiently providing services to their participants. We highly recommend for case managers to assess the participant status and use their judgment in making final decisions about what services should be provided. The recommendations should not replace participants being provided with basic need support, food and water, and continuous case management support.*

Our Dataset

Proprietary participant-level dataset

Originates from Homeless Management Information System (HMIS)[1]

- Not publicly available, with permission granted from a California Homelessness Organization for specific use in this project

CoC's spanning across Southern California each with 27 associated csv's

- Los Angeles
- Orange County
- San Diego
- Santa Barbara
- Santa Clara

Large dataset dependent on human data input

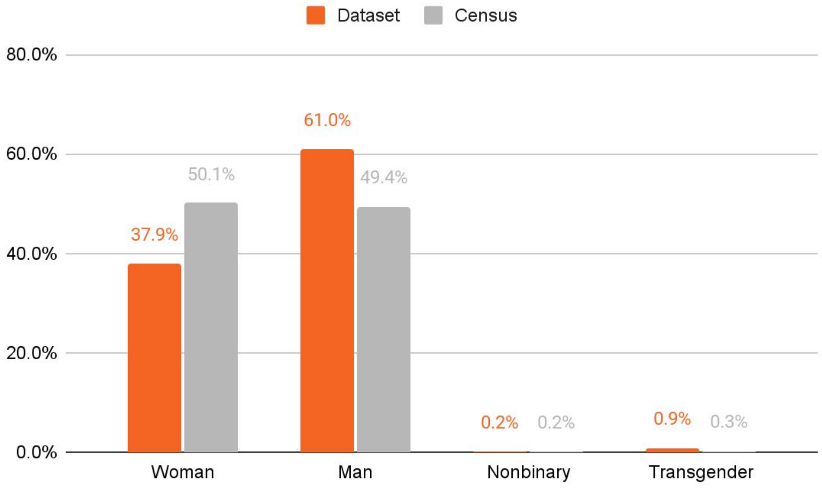
- Data entry reliant on manual case manager input
- Initial preprocessing resulted in around **400,000 records** to use for modeling

Data source: HMIS

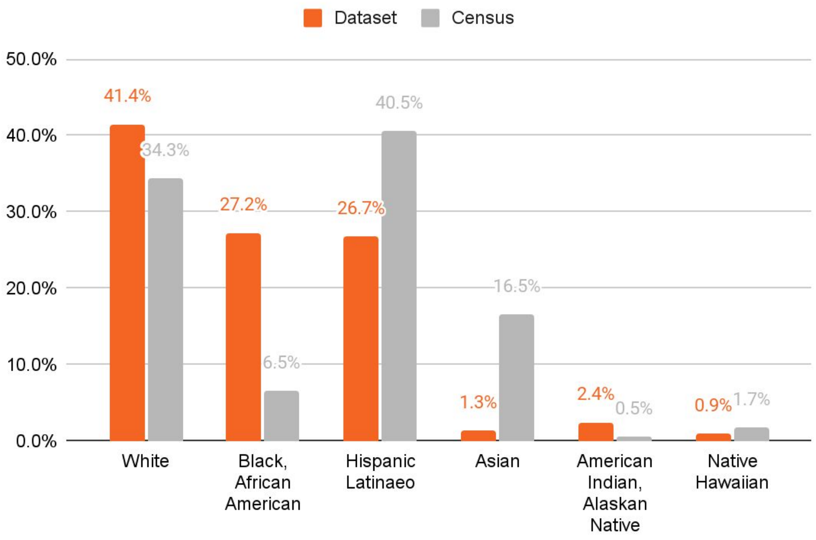
1. "[A] local information technology system used to collect client-level data and data on the provision of housing and services to individuals and families at risk of and experiencing homelessness". Link.

Population Demographics

Gender Distribution



Race Distribution



Data source: HMIS and [Census Data]

Key EDA at enrollment and exit

At Time of Enrollment

67.5% Of clients come from a place not meant for habitation

4.1% Of clients come from a temporary housing situation

1.8% Of clients come from a permanent housing situation

V.S.

At Time of Exit

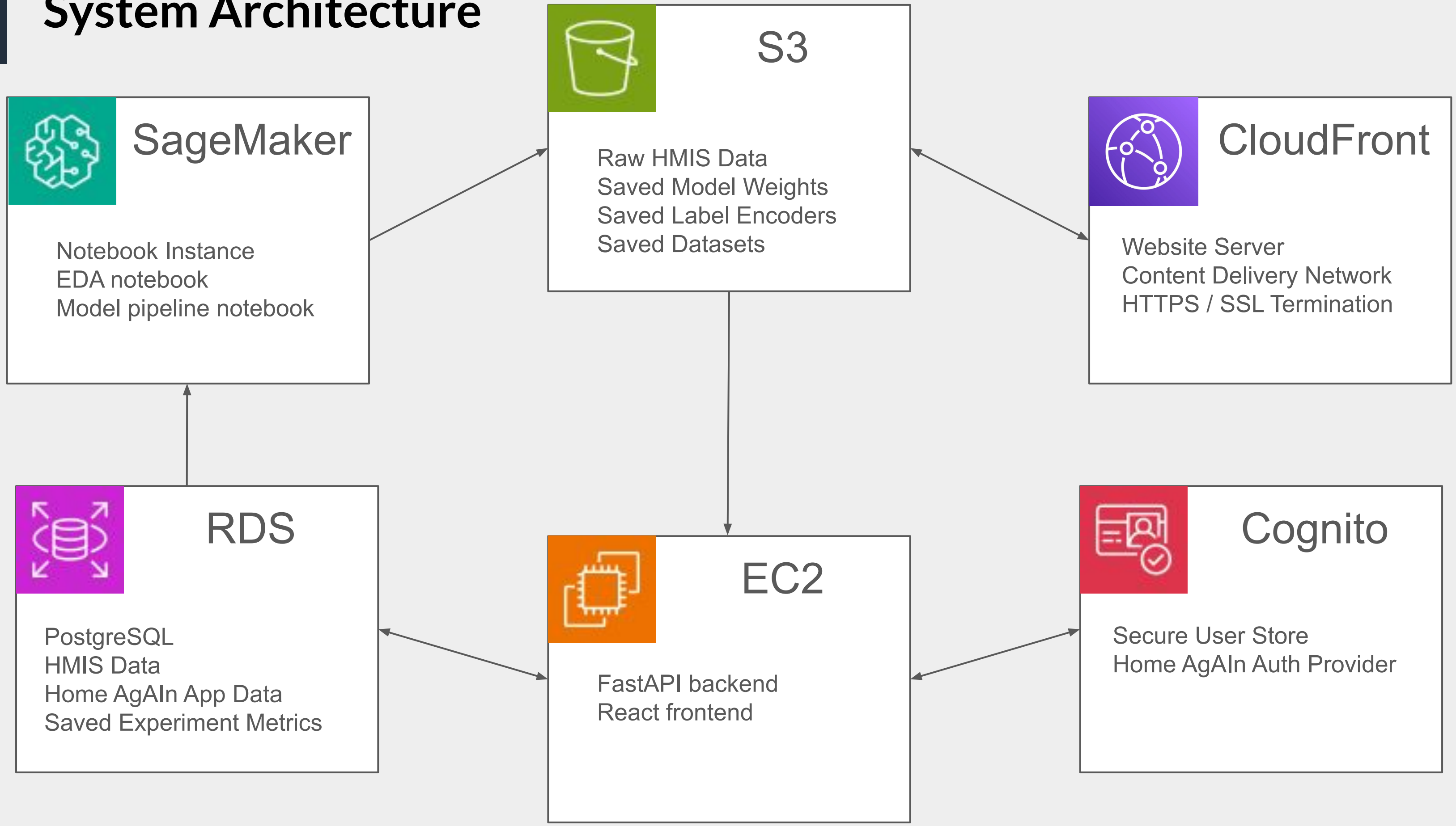
46.5% Of clients exit to a place not meant for habitation

3.1% Of clients exit to a temporary housing situation

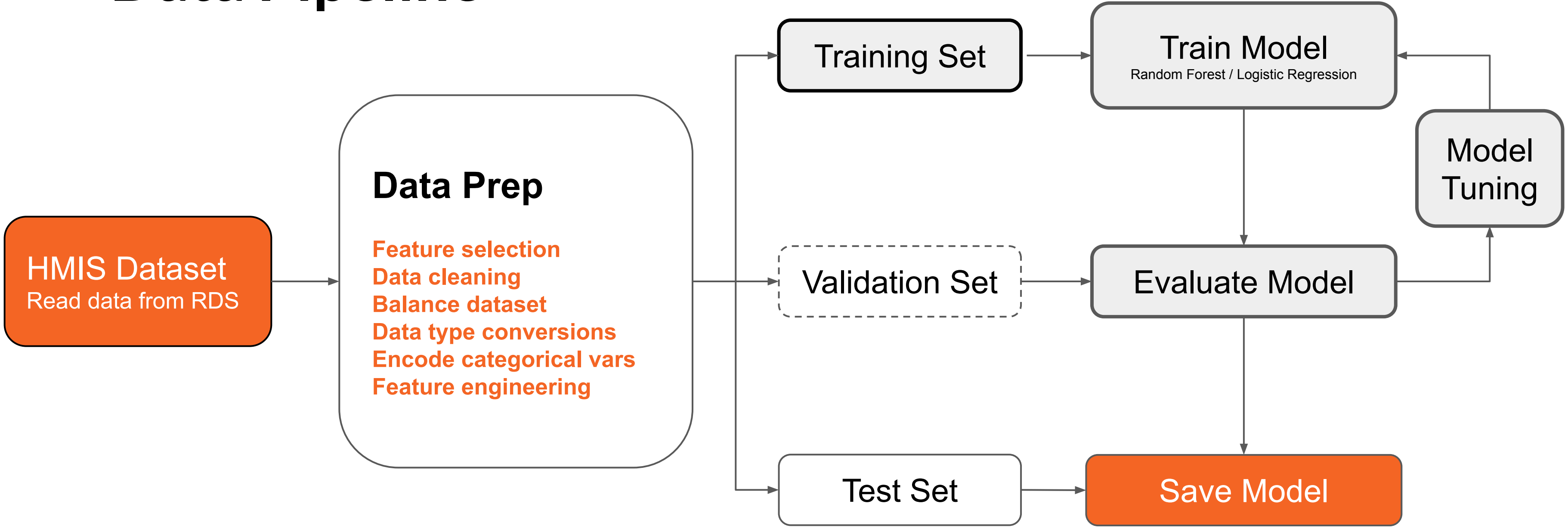
12.7% Of clients exit to a permanent housing situation

Data source: HMIS

System Architecture



Data Pipeline



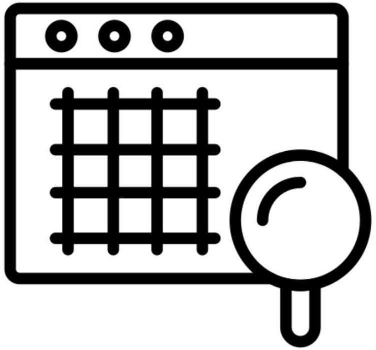
Robust and Flexible Pipeline Implementation

Key Classes:

DataSource	Queries data from RDS and encapsulates SQL
S3	Saves/Loads Models and Datasets via S3
FeatureEngineer	Implements feature engineering
Dataset	Implements train/test split
Model	Abstracts underlying model (eg. RF or LR)
Pipeline	**Orchestrates end-to-end Pipeline**

```python
class DataSource:
    def __init__(self, config):
        self.config = config
        self.connect()

    def connect(self):
        try:
            self.connection = psycopg2.connect(
                host=self.config.host,
                database=self.config.database,
                user=self.config.user,
                password=self.config.password,
                port=self.config.port
            )
            print("Connection to RDS successful")
        except Exception as e:
            print(f"Error connecting to RDS: {e}")
```

```python
class DataPipeline:
    def __init__(self, config):
        self.config = SimpleNamespace(**config)
        self.config.db = SimpleNamespace(**self.config.db)

    def run(self):
        run_start_time = datetime.now()

        data_source = DataSource(self.config.db)
        s3 = S3(bucket_name='capstone-hmis')

        dataset = s3.read_dataset(self.config.dataset_name)
```



Model V1 - Random Forest

Our first round of modeling

```
config = {
    "model_type": "RandomForest",
    "model_name": "random-forest-v10",
    "model_params": {
        'n_estimators': [50, 100, 200],
        'max_depth': [10, 20, None],
        'min_samples_split': [2, 5, 10],
        'min_samples_leaf': [1, 2, 4],
        'max_features': ['sqrt', 'log2', None]
    },
    "dataset_name": "v0.1.0",
    "dataset_max_size": None,
    "notes": "Full RandomForest Pipeline"
}
```

Strong performance

- Conducted extensive grid search across 243 hyperparameter specifications to fine tune
- Achieved high ROC AUC

Key challenges

- Model is not identifying causal relationships
- Service recommendations require "brute force"

Data source: HMIS

Model V2 - Logistic Regression

Causal Modeling with Interpretable Coefficients

Key Benefits

- Potential to identify **causal relationships** between services and permanent housing
- Coefficients that directly communicate value of different services
- Clear translation to **service recommendations**

Feature Engineering

- Implementation and achieving convergence required additional feature engineering
- **One-Hot** encoding and filling **null** values
- For simplicity, pivoted to a binary model (permanent housing outcome or not)

Data Imbalance

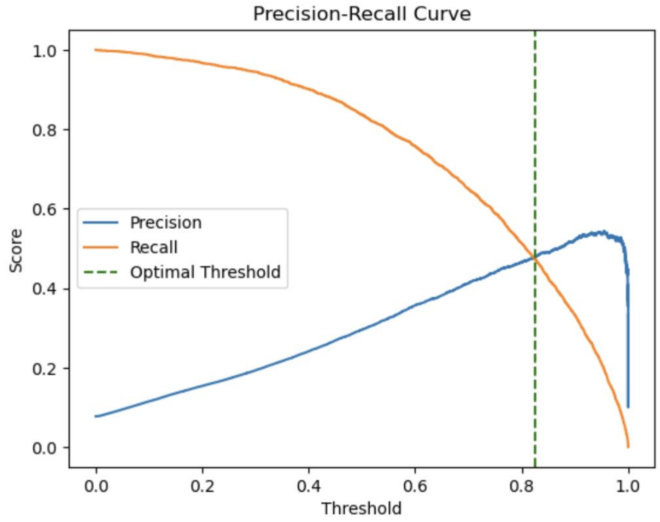
- Dataset is imbalanced with ~92% positive examples (permanent housing outcome) and ~8% negative examples[2]
- Utilized 'balanced' class weights to adjust for this imbalance[1]

Data source: HMIS
1. SKLearn Documentation Link, 2. From train dataset
Sources: SKLearn documentation/error. Professor guidance.

Logistic Regression Evaluation

Strong Recall with Moderate Precision

	Precision	Recall
Negative ***Non**-Permanent Housing Outcome*	0.98	0.83
Positive *Permanent Housing Outcome*	**0.30**	**0.84**



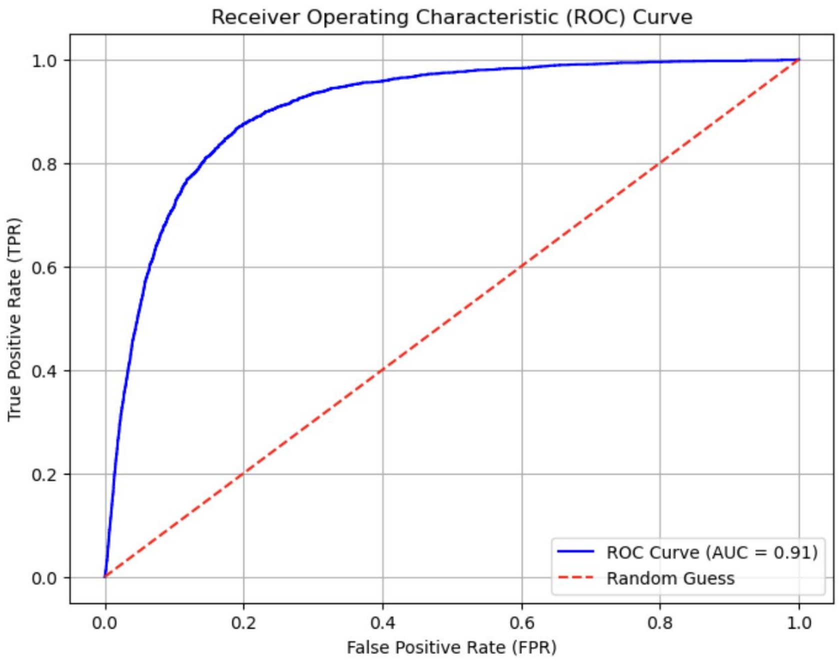
Model is "too optimistic"

- Model is identifying majority of the Permanent Housing Outcomes in the data (high recall)
- However, most examples model predicts as Permanent Housing Outcomes are false positives (low precision)

Data source: HMIS
Sources: ChatGPT, Wikipedia Link

Logistic Regression Evaluation

Good ROC AUC Results



AUC Performance is strong

- ROC curve for the model is compelling
- Overall AUC is only slightly below Random Forest

Data source: HMIS

Feature Importance

Coefficients and Odds Ratios of Key Features

	Coefficient	Odds Ratio *exp(coefficient)*[1]
Prior Living Situation - Permanent Housing	1.6	4.8
Education Support Service	1.2	3.2
Prior Living Situation - Temp Housing	1.0	2.8
Woman	0.8	2.2
Housing Search Service	0.7	2.0
Prior Living Situation - Homeless	-0.5	0.6
COVID Support Service	-0.7	0.5
Education - Data Not Collected	-2.0	0.1

Intuitive Results

- Those previously in permanent (or temporary housing) are more likely to exit to permanent housing
- Education and Housing Search services appear to improve odds of exiting to permanent housing
- Those entering a program already homeless are less likely to exit to permanent housing

Challenging Results - Correlation vs Causation

- Receiving COVID Support Services or not having data on Education may indicate a participant is in a challenging situation, not necessarily that these are directly causal

Data source: HMIS
Sources: UCLA Stats Link
1. UCLA Stats Link

Recommending Services to Improve Probabilities

- Logistic Regression Coefficients directly tell us which services will improve predicted probability of permanent housing
- For certain services, we have implemented heuristics (and added case management flags) to identify which participants may feasibly receive such services:

Home agAIn Demo

https://capstone-home-again.com/

Key Technical Takeaways

Rapid MVP Development

Rapid Development of End-to-End Solution Enabled by Robust System Architecture

Well-specified AWS architecture allowed for rapid connection of data pipeline, model results and front-end

Logistic Regression dramatically simplified service recommendation

Simple interpretation of coefficients eliminated any need to "brute force" and run multiple different service plans

Flexible Data Pipeline allowed rapid model switch

Pivot from Random Forest to Logistic Regression streamlined by pre-existing flexible data pipeline

Sources: May copy from AWS info session

Case Manager feedback on MVP

"This has been the first time I have felt some guidance towards supporting my participants without guessing what will help them the most"

Additional Asks: Participant profile picture, data quality checks

Image Source: [Link](#).

Improving causal modeling

- End-to-end model and application demonstrating enormous potential
- We still have plenty of opportunities to improve causal modeling robustness

Omitted Variable Bias (OVB)

	Coefficient	Odds Ratio
Substance Abuse Support Service	-0.14	0.87

Better control for individual characteristics to more accurately identify impact of different services

Feature Refinement

	Coefficient	Odds Ratio
Family Services	3.06	21.4

Split out reunification from parental support to better identify true driver of outsize impact

Data Source: HMIS

Identifying interactions

- Many services may be more or less helpful depending on the presence of different individual characteristics or other services
- Accounting for this could improve the customization of service recommendations and improve potential outcomes
- Some hypotheses to test:

Benefits - Disability Example[1]:

$$\text{logit}(p) = \ldots\ \beta_b{*}\textit{benefits\_assistance} + \beta_p{*}\textit{physical\_disability} + \boldsymbol{\beta_{bp}}{*}\boldsymbol{\textit{benefits\_assistance*physical\_disability}} + \ldots$$

Life Skills Service ↔ Participant Age

Hypothesis: Younger participants may benefit more from opportunities to gain life skills

Housing Search Service ↔ Move In and Rental Support

Hypothesis: Combining housing search with rental and move-in support may be synergistic

Benefits Assistance Service ↔ Physical Disability

Hypothesis: Disabled participants may be eligible for a wider range of benefits

Data source: HMIS
Sources: UCLA Stats Link
1. Copies from UCLA Stats Link

Enhance functionality and data collection

- Decompose the impact of each service recommendation to guide case managers in prioritization of recommended services

Service Recommendations

Permanent Housing Likelihood Improvement: **0.11099** (0.00378 → 0.11477)

- Collect additional participant characteristics information to address potential omitted variables

Duration of Substance Abuse (Years)	10
Incarcerated for Substance Abuse (Yes/No)	Yes

- Functionality requested by case manager tester

Participant Profile Picture

Data Quality Checks at time of Data Input

Data source: HMIS
Idea to decompose impact on probability may copy from UCLA Stats Link

Potential product roadmap

From MVP to adoption

Model and Functionality Enhancement → **Full QA Testing** → **Beta Testing** → **Full Rollout at First Homelessness Organization**

Model and Functionality Enhancement	Full QA Testing	Beta Testing	Full Rollout at First Homelessness Organization
Enhance MVP prior to testing	Professional QA testing to ensure robustness and usability	Beta testing with group of case manager "power users"	Launch with full case management team at an organization

“Data-driven service recommendations for the unhoused that empower case managers to bring participants Home Again”

Image Source: [Link]()

Appendix

Acknowledgements

Our team owes a California Homelessness Organization a huge thank you for their data and permission which allowed this project to happen. We would also like to thank the case manager who shared their insights on our MVP.

Additionally, we would like to thank our MIDS Capstone professors, Todd Holloway and Zona Kostic, for their invaluable guidance and feedback throughout the development of this project. We are also grateful to our MIDS Capstone classmates for their kind support and ideas throughout the project.

Additional Credits

- HMIS for data, data dictionaries, and data queries
- AWS for system components and deployment
- ChatGPT for coding support, help interpreting results, writing DB queries, and copy
- GitHub Copilot for coding support, website development, and copy
- SKLearn package and documentation for modeling and evaluation
- Prior MIDS courses for coding, modeling, statistics, web hosting, etc.
- Prior MIDS Capstone projects and website pages for inspiration and guidance