California Homelessness Challenge - Pocket Book Summary

Challenge Objective

Analyze homelessness trends across California counties using hospital encounters, demographics, and optionally fiscal funding, to:

- Build accurate forecasting models,
- Derive county-level vulnerability insights,
- Recommend data-backed policy interventions.

Datasets Provided

1. Homeless Demographics (Excel)
   - County-level data
   - Grouped by age, race, gender, etc.
   - Year granularity (2015-2021)
2. Hospital Encounters (CSV)
   - Homeless-related facility usage
   - Includes ownership type, shortage flags
   - County-level (2019-2020)
3. Fiscal Funding (Excel) (optional/bonus)
   - State-level homelessness funding (2015-2020)
   - May be used for trend correlation

Modeling Target

Predict TOTAL_HOMELESS for each California county, based on available features (hospital usage, shortages, demographics, etc.)

Tools & Methods Expected

- Pandas / NumPy / scikit-learn / matplotlib / seaborn
- Advanced models: Random Forest, XGBoost, Gradient Boosting
- Dimensionality Reduction: PCA
- Clustering: KMeans or other
- Forecasting: Lag features + prediction horizon

- What-if scenario simulation

Final Question Set (15 Total)

Data Prep & Merging (Q1-Q3)

- Standardize columns, handle missing data
- Merge datasets by County, prepare for modeling

EDA & Trends (Q4-Q6)

- Visualize county-level homelessness
- Compare hospital usage per homeless individual
- Explore variable relationships

Feature Engineering (Q7-Q9)

- Ratios, lags, encodings
- Rolling averages, YOY changes
- Cluster counties by vulnerability

Modeling & Optimization (Q10-Q12)

- Baseline linear models
- Tree-based + XGBoost models
- PCA or cluster-based models

Forecasting & Policy (Q13-Q15)

- Forecast future homeless counts
- Simulate a policy shift & evaluate
- Recommend 3 high-need counties with justification

Evaluation Rubric

Category	Points

California Homelessness Challenge - Pocket Book Summary

Data Preparation	100 pts
Exploratory Analysis	150 pts
Feature Engineering	150 pts
Baseline Modeling	100 pts
Advanced ML (Tuning, Ensemble)	200 pts
Forecasting + Strategy (Bonus)	100 pts
Policy Recommendations	100 pts
Notebook Quality & Clarity	100 pts

Submission Guidelines

- All work must be inside a single, clean notebook.
- Include clear comments, section headings, and insightful interpretations.
- No dashboards, no external reports.
- You may use external data if cited, but core insights must rely on the provided datasets.

Tips for Teams

- Prioritize clarity over complexity. Judges will read your code.
- Focus on interpretability - explain why features matter.
- Don't overfit - ensure your model generalizes to unseen counties.
- Visuals matter! Make plots readable and label them well.
- Document assumptions and challenges clearly.