PREDICTING HOUSING PRICES BASED ON JOB FACTORS

Understanding the Link Between Employment and Real Estate in California

Presenter: Donn Bryan Julian
Date: 10/15/2024

PROJECT OVERVIEW

- The objective of this project was to predict housing prices in California based on job-related factors.
- We aimed to understand how employment opportunities, income levels, and housing characteristics influence the affordability of homes.
- The analysis combines housing data with career data to uncover key insights for policymakers, real estate developers, and the general public.

DATASETS USED

- Housing Data: Median age, rooms, bedrooms, population, households, median income, house value, and location.
- Job Data: Job listings, job titles, salary ranges, work type, and locations (focused on Alameda and Contra Costa counties).
- By merging these datasets, we gained insights into the relationship between employment opportunities and housing affordability.

DATA CLEANING STEPS

- Filtered job data to focus on Alameda and Contra Costa counties for localized analysis.
- Removed duplicate job listings to ensure accuracy.
- Created a unified "Location" column in the housing data to align with the job dataset.
- Extracted salary range values into "min_salary" and "max_salary" columns for analysis.

ANALYTICAL APPROACH

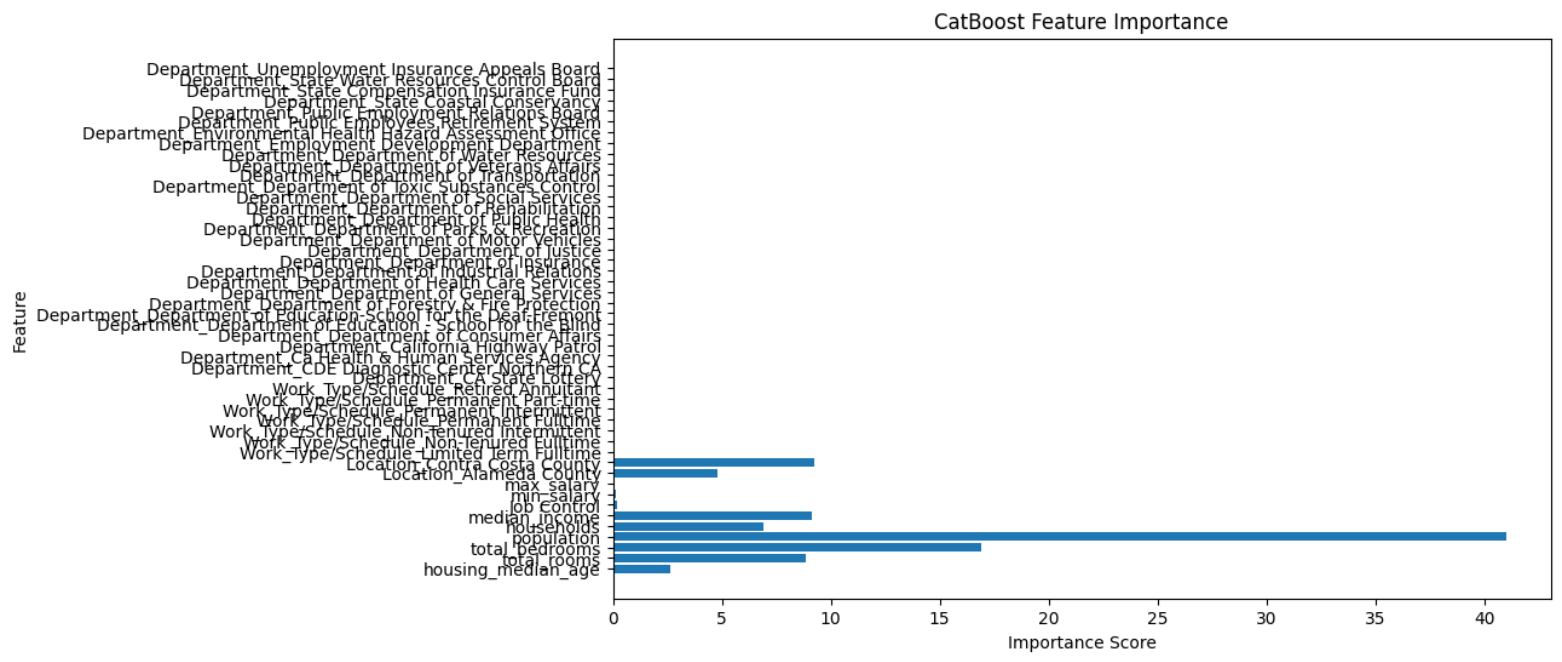
- One-Hot Encoding was applied to categorical features like job types and department names for machine learning purposes.
- We experimented with various regression models, including Linear Regression, Random Forest, Gradient Boosting, and CatBoost, to predict median house values.
- Feature importance analysis was used to understand the significance of each predictor.

MODEL RESULTS

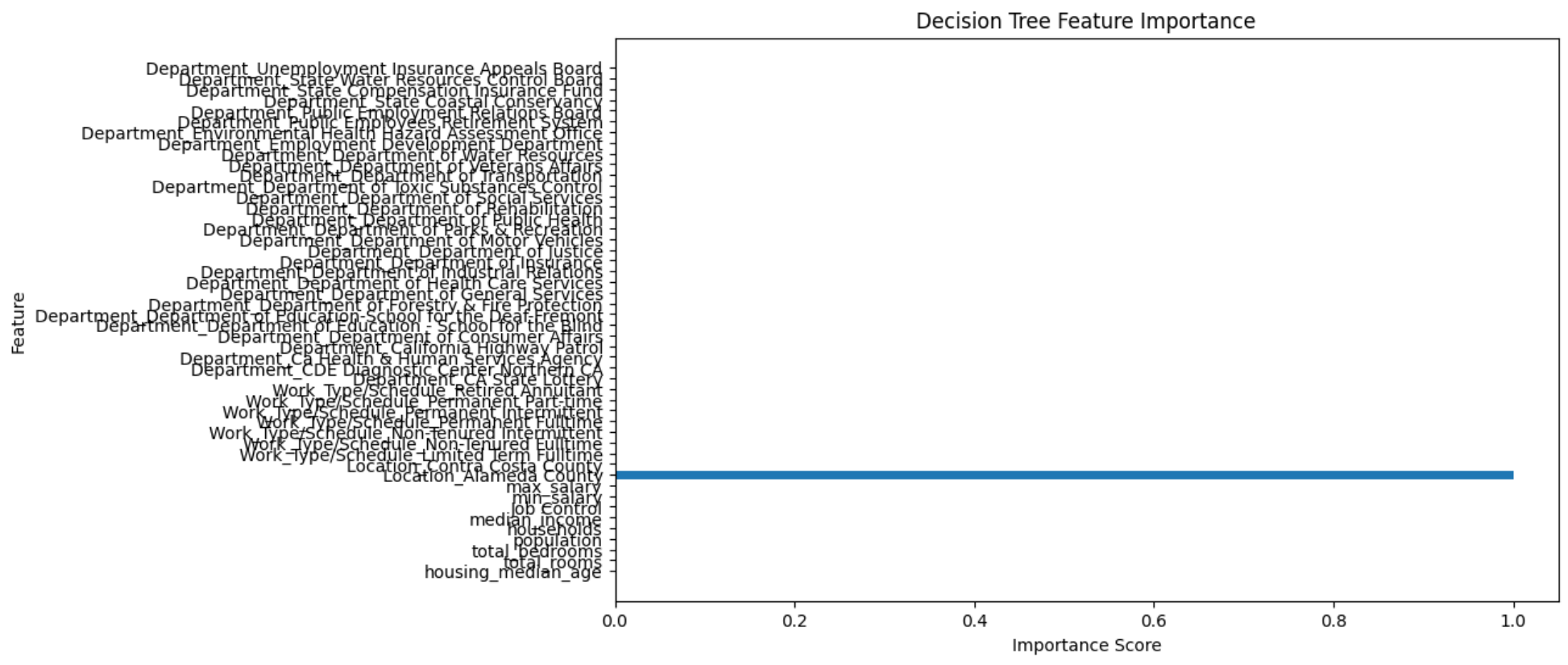
- Several models showed near-perfect R^2 scores, indicating strong predictability, but this was largely influenced by data complexity.
- Reinforcement Learning techniques were used to optimize model blending for higher accuracy, resulting in an R^2 of 0.91 during cross-validation.

KEY DRIVERS OF HOUSING PRICES

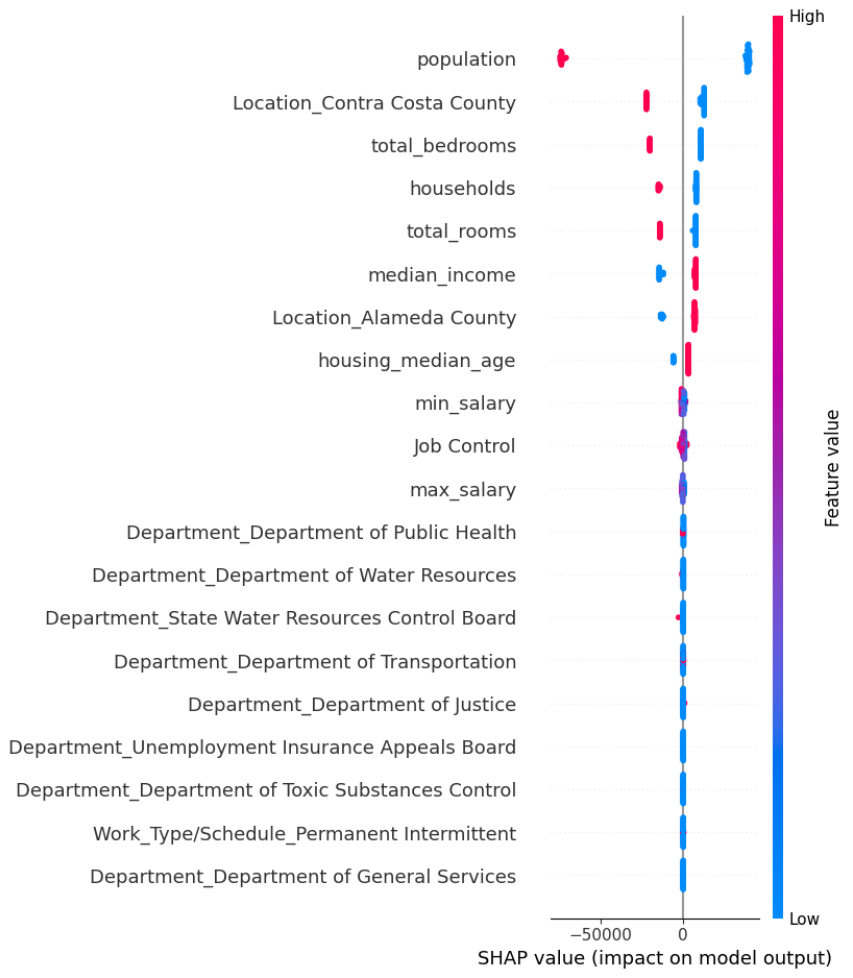
- CatBoost Model Insights: Median Income and Population were highly significant predictors.



- The Decision Tree model identified Households as the most significant feature in predicting housing prices. This reflects the Decision Tree's simple splitting mechanism, focusing on features that provide clear distinctions in the data. However, the model's limited complexity may overlook subtle relationships between other features and housing prices.



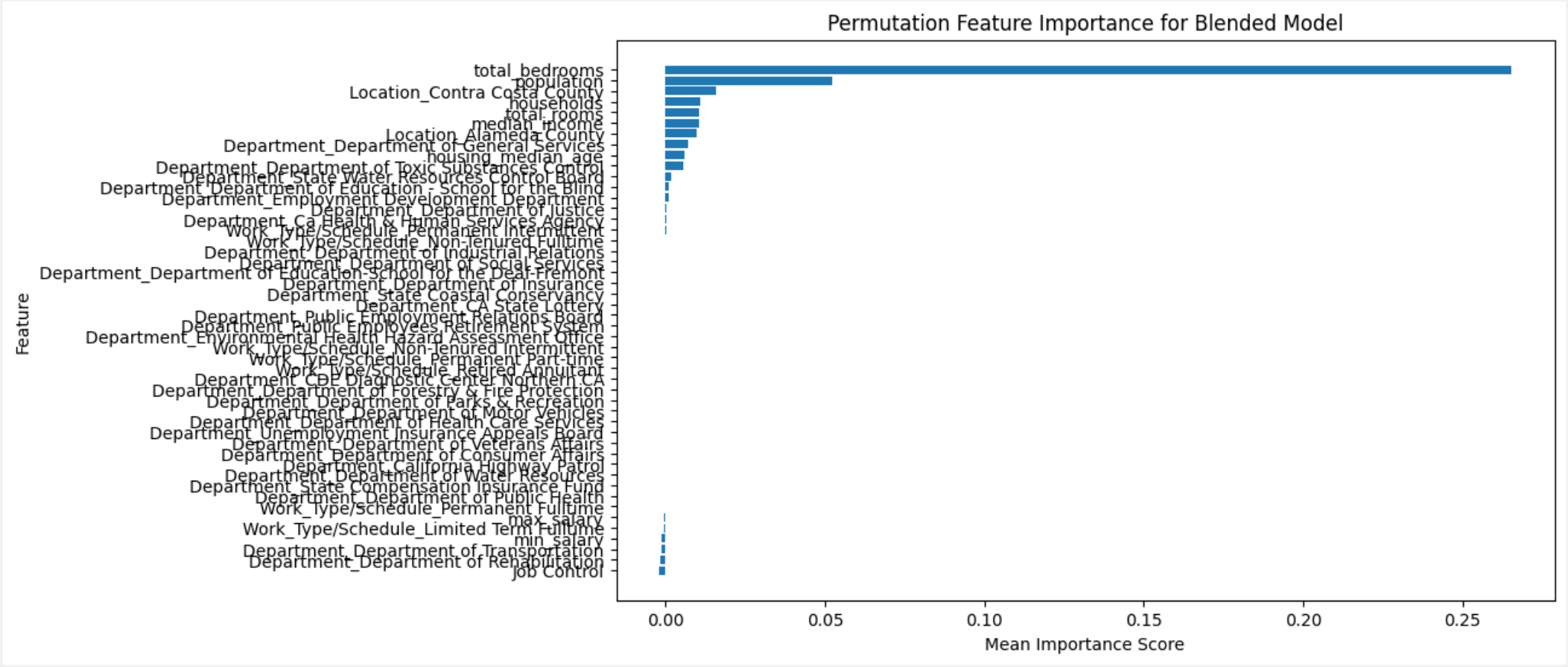
SHAP ANALYSIS



SHAP Analysis:
The SHAP summary plot highlighted the influence of features like Population, Location, and Total Bedrooms on housing prices, providing deeper interpretability of the model's decisions.

FINDINGS

- Economic factors like Median Income play a key role in determining housing affordability.
- Job-related factors like location (e.g., Contra Costa County) also have a measurable impact.
- Population and the number of available rooms are significant drivers of home values, potentially indicating market demand.


Permutation Feature Importance for Blended Model
Feature
total_bedrooms
population
Location_Contra Costa County
households
total_rooms
median_income
Location_Alameda County
Department_Department of General Services
housing_median_age
Department_Department of Toxic Substances Control
Department_State Water Resources Control Board
Department_Department of Education - School for the Blind
Department_Employment Development Department
Department_Department of Justice
Department_Ca Health & Human Services Agency
Work_Type/Schedule_Permanent Intermittent
Work_Type/Schedule_Non-Tenured Fulltime
Department_Department of Industrial Relations
Department_Department of Social Services
Department_Department of Education-School for the Deaf-Fremont
Department_Department of Insurance
Department_State Coastal Conservancy
Department_CA State Lottery
Department_Public Employment Relations Board
Department_Public Employees Retirement System
Department_Environmental Health Hazard Assessment Office
Work_Type/Schedule_Non-Tenured Intermittent
Work_Type/Schedule_Permanent Part-time
Work_Type/Schedule_Retired Annuitant
Department_CDE Diagnostic Center Northern CA
Department_Department of Forestry & Fire Protection
Department_Department of Parks & Recreation
Department_Department of Motor Vehicles
Department_Department of Health Care Services
Department_Unemployment Insurance Appeals Board
Department_Department of Veterans Affairs
Department_Department of Consumer Affairs
Department_California Highway Patrol
Department_Department of Water Resources
Department_State Compensation Insurance Fund
Department_Department of Public Health
Work_Type/Schedule_Permanent Fulltime
max_salary
Work_Type/Schedule_Limited Term Fulltime
min_salary
Department_Department of Transportation
Department_Department of Rehabilitation
Job Control
0.00
0.05
0.10
0.15
0.20
0.25
Mean Importance Score

PROJECT CHALLENGES

- The dataset was limited to Alameda and Contra Costa counties, restricting broader generalizability.
- Potential data quality issues, such as missing values and duplicates, may have influenced model predictions.
- Overfitting concerns, especially with simpler models like Decision Trees.

FINAL TAKEAWAYS

- Housing prices in California are highly influenced by a combination of economic and employment-related factors.
- Predictive models can help inform decision-making for homebuyers, policymakers, and urban planners.
- This work emphasizes the need for more integrated data to drive decisions about housing and economic development.

LOOKING AHEAD

- Expand analysis to other counties in California for a more comprehensive view.
- Incorporate more job characteristics, such as job satisfaction or commuting distances, to enrich the analysis.
- Explore other machine learning models or deeper neural networks for potentially improved predictions.

CLOSING REMARKS

- This project provided valuable insights into the link between job factors and housing affordability.
- The findings can contribute to informed decisions in housing policy, urban planning, and economic development.
- Thank you for your attention, and we welcome any questions or thoughts on how we can further expand this analysis.

QUESTIONS?