RENTAL DEMAND ANALYSIS & SYSTEM MODERNISATION

Presented by: Nicolette Woolery

Project Summary

This project combines machine learning, database design, and ethical computing to guide business decisions and support the company’s digital transformation amid a shifting rental market.

Client Brief

Project Overview

- Objective: Investigate the decline in rental housing demand and propose solutions.
- Business risk: under-occupied listings and lost revenue.
- Goal: Identify demand drivers, modernise systems, ensure compliance.
- Focus: Data-driven insights, system redesign, and privacy-aware client engagement.

Project Objectives

•Analyze characteristics that influence rental demand.

•Key Questions:

- What predicts low or high demand?
- Is there a correlation with rent or property type?
- What size properties generate the most demand?

Project Tasks

1. Analyze demand patterns using machine learning (WEKA).
2. Replace CSV storage with a normalized relational database.
3. Develop scalable infrastructure for future global operations.
4. Address privacy issues for a proposed public-facing application.

Rental Demand Investigation Deliverables

- Identify property attributes that influence demand.
- Assess correlations with rent, type, and square footage.
- Build classifiers/regression models to find patterns.
- Deliver insights using visualisations and performance metrics.

Data Storage & Scalability

- Design relational database (ER diagram, 3NF).
- Provide SQL to insert and query data.

- Recommend distributed infrastructure (cloud, message brokers).
- Justify based on volume, responsiveness, and scaling needs.

Data Storage & Scalability

- Build user platform for property browsing and recommendations.
- Identify 3 privacy risks:
  - Data protection
  - Secure handling
  - Ethical analytics
- Recommend strategies to mitigate each issue.

Project Solution

Data Storage & Scalability

- Cleaned and preprocessed CSV using Python.
- Removed irrelevant columns.
- One-hot encoding applied.
- Converted datasets into ARFF for WEKA.

Analysis Techniques

- Decision Tree (J48) for demand classification - Classifies low/high demand using discrete variables
- Pearson Correlation for rent/type vs demand - Assesses impact of rent/type on demand.
- Linear Regression + Decision Tree to analyze 'sqfeet' effect - Evaluates relationship between sqfeet and demand

Justification of Techniques

- J48 is suitable for categorical insights and interpretability.
- Pearson captures linear relationships.
- Combined regression and decision tree detect complex patterns.

Key Insights

- Bedrooms, laundry, pet policies strongly affect demand.
- Weak correlation between rent and demand.
- Optimal sqfeet range for high demand: ≤ 3235 sqft.

Database Approach

- Designed relational database in 3NF.
- Used Property, Location, Policy, Amenity tables.
- Maintained referential integrity and minimized redundancy.

SQL Examples

- Insert new listing
- Query properties: rent ≤ 1000, pets allowed, CA
- Avg rent per state

Technology Justification

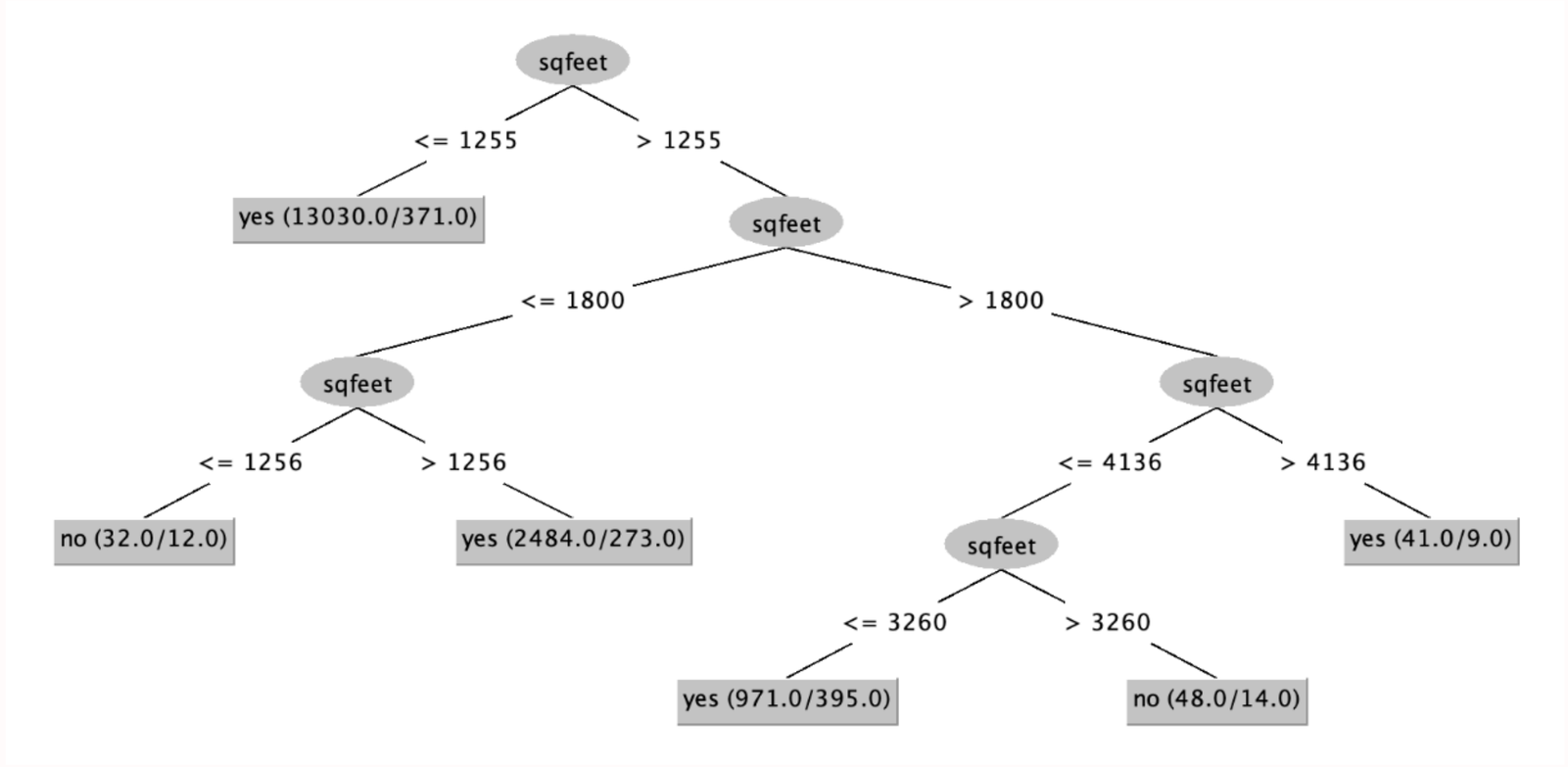
- RDBMS chosen for structure, integrity, query capabilities.
- Compared RDBMS, NoSQL, and Data Warehousing.
- RDBMS balances operational efficiency and analysis.

Privacy & Compliance

- Consent & Data Transparency: Clear opt-in policies and explain data usage.
- Data Retention: Auto-deletion policies - Define and enforce retention schedules.
- Data Security: Encryption, access control, regular audits.

Results Summary

- Influential Variables: bedrooms, pets, laundry.
- Rent/type show weak correlation to demand.
- Optimal property size for high demand: up to 3235 sqft.



Square Feet and Demand Decision Tree Model and Visualisation

What Predicts Low or High Demand?

Top Predictors (from J48 Decision Tree):

- Bedrooms: Properties with 1–2 bedrooms showed higher demand than larger ones.
- Laundry Options: In-unit laundry facilities significantly increase demand.
- Pet Policies: Listings allowing both cats and dogs had higher interest.
- Smoking Policy: Non-smoking-friendly properties tended to rank higher in demand.
- Parking: Reserved or attached parking options also contributed positively.

These features were consistently used in splits by the decision tree to predict “low” vs. “high” demand classes.

Is There a Correlation Between Rent or Property Type and Demand?

Pearson Correlation Results:

• Rent vs Demand: Weak negative correlation (-0.18), meaning that higher rent slightly reduces demand, but not significantly enough to be a strong predictor.

• Type vs Demand: Minimal correlation; property type alone does not determine demand. However, when paired with other features (like size or amenities), it plays a supporting role.

Conclusion: Rent and type are not strong standalone predictors. Amenities and living preferences matter more.

What Size Properties Generate the Most Demand?

Sqfeet vs Demand Analysis (Regression + J48):

• Properties sized between 500 – 3235 square feet had the highest likelihood of being marked as high demand.

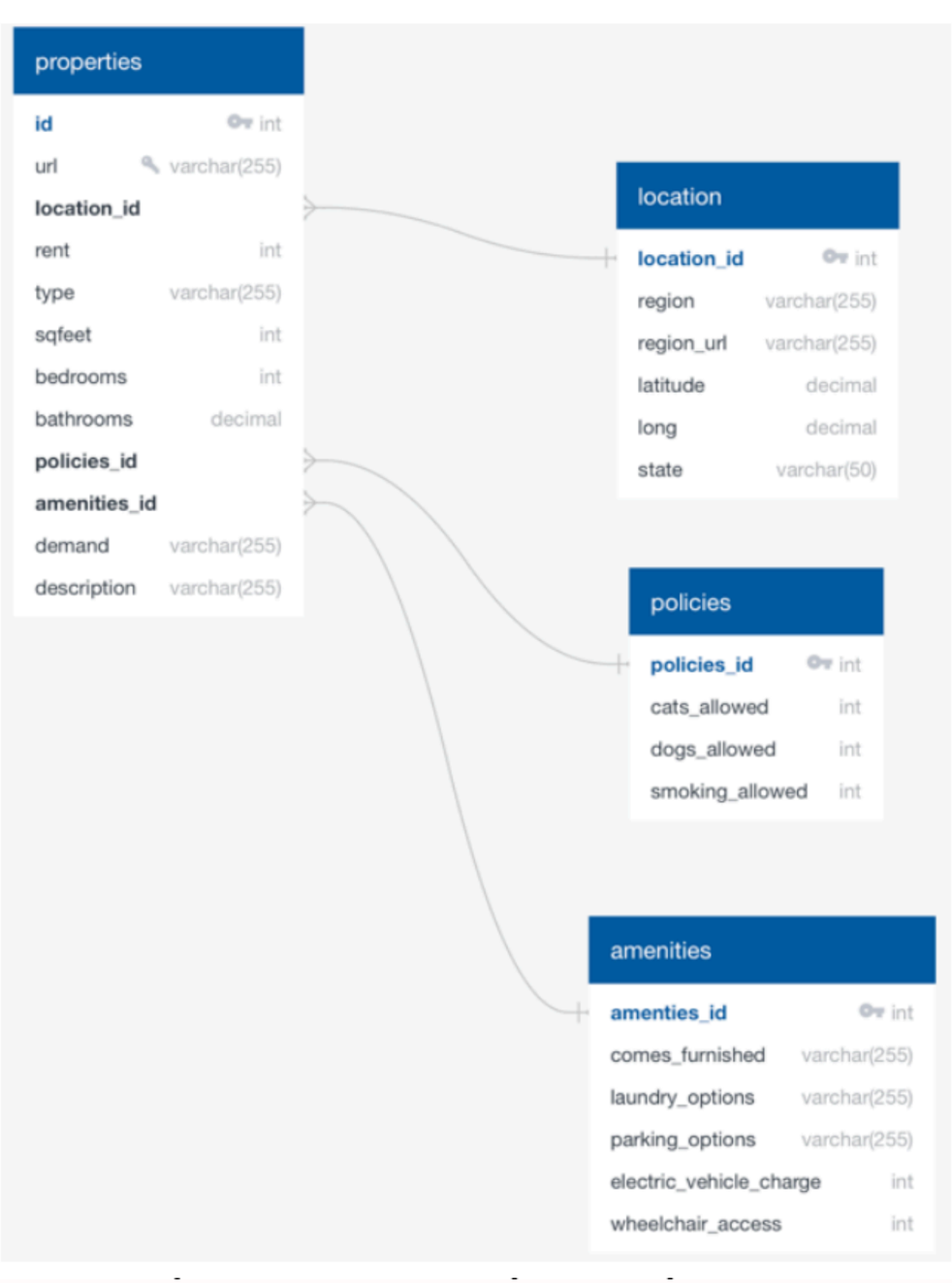
• Extremely large properties (>3500 sqft) often showed lower demand, likely due to higher rent and limited target audience.

Database Design

•Built normalised schema in 3NF.

• Key tables: Property, Location, Amenities, Policy.

• Ensures consistency, scalability, and efficiency.



SQL Functionality Examples

- INSERT: Add a new property listing.
- SELECT: Filter by rent ≤ 1000, pets allowed, and state = 'ca'.
- AGGREGATE: Calculate average rent by state.

```
CREATE TABLE Property (
    id INT PRIMARY KEY,
    url VARCHAR(255) UNIQUE,
    location id INT,
    rent INT,
    type VARCHAR(255),
    sqfeet INT,
    bedrooms INT,
    bathrooms DECIMAL(3,1),
    policy_id INT,
    amenity_id INT,
    deman VARCHAR(255),
    description VARCHAR(255),
    FOREIGN KEY (location_id) REFERENCES Location(location_id),
    FOREIGN KEY (policy_id) REFERENCES Policy(policy_id),
    FOREIGN KEY (amenity_id) REFERENCES Amenity(amenity_id)
);

CREATE TABLE Amenity (
    amenity_id INT PRIMARY KEY,
    comes_furnished BOOLEAN,
    laundry_options VARCHAR(255),
    parking_options VARCHAR(255),
    electric_vehicle_charge BOOLEAN,
    wheelchair_access BOOLEAN,
);

CREATE TABLE Policy (
    policy_id INT PRIMARY KEY,
    cats_allowed BOOLEAN,
    dogs_allowed BOOLEAN,
    smoking_allowed BOOLEAN,
);

CREATE TABLE Location (
    location_id INT PRIMARY KEY,
    region VARCHAR(255),
    region_url VARCHAR(255),
    latitude DECIMAL(9,6),
    state VARCHAR(50)
);
```

SQL Statement for creating tables

Scalable Solutions

- Scenario: International expansion, higher data volume.
- Solution: Distributed architecture using cloud technologies.
- Tools: Hadoop, Apache Spark, or AWS Redshift.
- Enables rapid analytics and real-time alerts.

Technology Stack

Technology	Purpose
WEKA (v3.8.5)	Primary data mining and machine learning tool. Used for classification (J48), regression, and correlation analysis.
Python (Pandas, NumPy)	Data cleaning, transformation, and preprocessing of the raw CSV data before importing into WEKA.
Spreadsheet Software (Excel/Google Sheets)	Initial data exploration, formatting, and identifying anomalies.
.ARFF Format	Required file format for WEKA. Converted from cleaned CSV using Python.
SQL	Used to write queries and demonstrate relational database operations such as inserts, filters, and aggregates.
ERD Tools (e.g., Draw.io, Lucidchart, Plant UML)	To design and present the ER diagram (3NF) for the proposed relational database.

Project Link

nixxwoolery/
RentalDemandAnalysis

Rental Demand Analysis & System Modernisation

1 Contributor | 0 Issues | 0 Stars | 0 Forks

nixxwoolery/RentalDemandAnalysis: Rental Demand Analysis & System Modernisation

Rental Demand Analysis & System Modernisation. Contribute to nixxwoolery/RentalDemandAnalysis development by creating an account on GitHub.

GitHub

Final Recommendations

- Use machine learning insights to inform listings strategy.
- Transition from flat files to a relational database.
- Prepare for scale with distributed architecture.
- Build a secure, user-friendly public platform.

Summary

- High demand is driven by liveability features: pets allowed, laundry, and modest size.
- Low demand is linked to higher rent, excessive size, and lack of in-unit amenities.
- These insights can guide marketing, pricing, and listing strategy for future properties.

Conclusion

A combined machine learning and system design approach helps uncover demand trends, ensures scalable data storage, and addresses user privacy in a future web-based platform.