Toronto AirBnB Investment Navigator - Project Outline

# Project Overview

Our team will develop the **Toronto AirBnB Investment Navigator**, a data-driven tool that helps real estate investors identify high-potential neighborhoods for short-term rental investments in Toronto. By analyzing Inside AirBnB data and integrating property values, we'll provide actionable insights on neighborhood investment potential and expected returns for the Toronto short-term rental market.

# Team Members

* **Sean Schallberger**
* **Bryan Carney**
* **Jitesh Makan**

# Primary Stakeholders: Real Estate Investors

Our tool is designed specifically for:

* **Individual property investors** seeking to purchase AirBnB properties
* **Current AirBnB hosts** looking to expand their portfolio

Our stakeholders need:

* Clear identification of high-ROI neighborhoods
* Understanding of property value to rental income ratios
* Data-driven investment recommendations

# Data Sources

### Primary Data (MVP)

1. **Inside AirBnB Toronto** (March 2, 2025):
   * listings.csv.gz - Property details, pricing, and features
   * calendar.csv.gz - Availability and booking data
   * reviews.csv.gz - Guest feedback and ratings
   * neighbourhoods.geojson - Geographic boundaries
2. **Toronto Property Values**:
   * Toronto Real Estate Board neighborhood price data

## Machine Learning Approach

Our analysis will use a simplified set of models to provide clear investment insights:

|  |  |  |
| --- | --- | --- |
| Investor Need | ML Model | Output |
| Identify investment-worthy neighborhoods | **K-means Clustering** | Neighborhood categories by investment potential |
| Understand rental performance | **Linear Regression** | Key factors affecting rental income |
| Compare investment options | **Investment Score Algorithm** | ROI ranking for neighborhoods |

### Model Details

1. **K-means Clustering**
   * **Purpose**: Group neighborhoods by AirBnB performance
   * **Features**: Average price, occupancy rate, review scores
   * **Output**: Investment potential categories (High, Medium, Low)
2. **Linear Regression**
   * **Purpose**: Identify factors that influence rental income
   * **Features**: Property size, amenities, location
   * **Output**: Feature importance for rental success
3. **Investment Score Algorithm**
   * **Purpose**: Rank neighborhoods by ROI potential
   * **Formula**: (Avg. Revenue / Property Value) × Occupancy Rate

# Simplified Project Timeline (14 Days)

### Pre-Project Setup

* Download all datasets from Inside AirBnB Toronto
* Acquire Toronto property value data

#### Sprint 1: Data Cleaning & Exploration (5 days)

**Tasks**:

* **All Team**: Jointly clean and explore datasets during Monday class
* **Sean**: Focus on listings data analysis
* **Bryan**: Handle property value integration
* **Jitesh**: Analyze booking and review patterns

**Checkpoint (Friday)**: Complete data cleaning and initial analysis

#### Sprint 2: Modeling & Visualization (5 days)

**Tasks**:

* **Sean**: Implement K-means clustering
* **Bryan**: Develop linear regression and ROI calculation
* **Jitesh**: Create interactive map visualization

**Checkpoint (Friday)**: Complete core models and visualization framework

#### Sprint 3: Integration & Presentation (4 days)

**Tasks**:

* **All Team**: Finalize dashboard and documentation
* **All Team**: Prepare and practice presentation

**Final Presentation**: Deliver project presentation

# Exploratory Data Analysis Plan

Our focused EDA will examine:

1. **Neighborhood Analysis**
   * Property distribution across Toronto
   * Price ranges by neighborhood
   * Review ratings geographical patterns
2. **Financial Analysis**
   * Occupancy rates calculation
   * Revenue potential estimation
   * Property value to rental income ratios

# Technical Implementation

### Data Processing Simplification

1. **Initial Cleaning**:
   * Handle missing values and remove duplicates
   * Focus on essential fields for analysis
2. **Core Feature Creation**:
   * Calculate occupancy rates
   * Estimate monthly revenue
   * Compute basic ROI metrics

### Focused Machine Learning

1. **Neighborhood Clustering**:
   * Group similar neighborhoods (3-5 clusters)
   * Assign investment potential categories
2. **Basic Linear Model**:
   * Identify key revenue drivers
   * Calculate ROI by neighborhood

### Streamlined Visualization

1. **Interactive Map**:
   * Color-coded neighborhoods by investment potential
   * Simple popup with key metrics
2. **Investment Dashboard**:
   * Top 10 neighborhoods by ROI
   * Basic comparison charts

## Repository Structure

|  |
| --- |
| toronto\_airbnb\_investment\_navigator/  │  ├── data/  │ ├── raw/ # Original datasets  │ └── processed/ # Cleaned datasets  │  ├── notebooks/  │ ├── 1\_data\_cleaning.ipynb  │ ├── 2\_exploratory\_analysis.ipynb  │ ├── 3\_modeling.ipynb  │ └── 4\_visualization.ipynb  │  ├── src/  │ ├── data\_processing.py # Data cleaning functions  │ ├── modeling.py # ML model implementation  │ └── visualization.py # Map code  │  ├── app/  │ ├── static/ # CSS and JavaScript  │ └── templates/ # HTML files  │  ├── results/ # Output files and visualizations  │  ├── individual\_work/ # Individual team member work  │ ├── sean/ # Sean's individual work files  │ ├── bryan/ # Bryan's individual work files  │ └── jitesh/ # Jitesh's individual work files  │  ├── requirements.txt  └── README.md |

## Conclusion

The Toronto AirBnB Investment Navigator will provide real estate investors with clear, actionable insights to guide their short-term rental investment decisions. By focusing on essential analysis and a simplified approach, we'll ensure a high-quality, stakeholder-friendly product that meets all project requirements while accommodating our team's varying schedules and technical backgrounds.