Toronto AirBnB Pricing Assistant (TAPA)

# Project Outline & Timeline

### Project Overview

Our team will develop the **Toronto AirBnB Pricing Assistant & Profit Maximizer (TAPAM)**, a data-driven tool that helps AirBnB hosts and property investors optimize their rental pricing strategy in Toronto. By analyzing Inside AirBnB data and identifying key amenity value drivers, we'll provide actionable insights on optimal pricing, amenity recommendations, and stay duration strategy for the Toronto short-term rental market.

# Team Members

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

# Primary Stakeholders: AirBnB Hosts and Property Investors

Our tool is designed specifically for:

* **Current AirBnB hosts** looking to optimize pricing and amenities
* **Potential investors** evaluating rental strategies
* **Property owners** considering conversion to AirBnB rentals

Our stakeholders need:

* Data-driven price recommendations
* Understanding of amenity value impact
* Guidance on short-term vs. long-term stay strategy

# 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

# Machine Learning Approach

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

|  |  |  |
| --- | --- | --- |
| Host/Investor Need | ML Model | Output |
| Optimize property pricing | **Linear Regression** | Recommended price range |
| Identify high-value amenities | **Feature Importance Analysis** | Amenity value ranking |
| Choose stay duration strategy | **Comparative Analysis** | Short vs. long-term recommendation |

### Model Details

1. **Linear Regression (Price Prediction)**
   * **Purpose**: Determine optimal pricing for properties
   * **Features**: Property type, location, amenities, accommodations
   * **Output**: Recommended price range
2. **Feature Importance Analysis**
   * **Purpose**: Identify amenities that increase property value
   * **Features**: Binary amenity indicators
   * **Output**: Ranked list of price-boosting amenities
3. **Stay Duration Analysis**
   * **Purpose**: Recommend short vs. long-term stay strategy
   * **Method**: Comparative neighborhood performance analysis
   * **Output**: Stay length recommendation by property type and location

# Simplified Project Timeline (14 Days)

### Pre-Project Setup

* Download all datasets from Inside AirBnB Toronto
* Perform initial data exploration and cleaning

#### Week 1: Data Engineering & Model Development (7 days)

**Days 1-2: Project Setup & Data Cleaning**

* Create GitHub repository with proper structure
* Standardize property types to 5 key categories
* Extract and categorize amenities
* Handle missing values in essential fields

**Days 3-4: Feature Engineering & Initial Modeling**

* Create binary amenity features
* Implement stay length classification
* Begin developing price prediction model
* Analyze amenity patterns

**Days 5-7: Model Development & Optimization**

* Finalize price prediction model
* Extract amenity importance values
* Develop stay length recommendation logic
* Begin documentation and visualization

#### Week 2: Implementation & Presentation (7 days)

**Days 8-9: Analysis & Visualization**

* Create key visualizations
* Implement recommendation function
* Document model performance
* Begin integration of components

**Days 10-12: Testing & Documentation**

* Test with various property scenarios
* Complete GitHub documentation
* Prepare presentation materials
* Create demonstration examples

**Days 13-14: Finalization & Presentation**

* Finalize all components
* Complete presentation slides
* Practice presentation
* Submit final project

# Exploratory Data Analysis Plan

Our focused EDA will examine:

1. **Property Analysis**
   * Distribution of property types across Toronto
   * Price ranges by property category and neighborhood
   * Amenity prevalence and co-occurrence patterns
2. **Stay Duration Analysis**
   * Comparison of short vs. long-term rental properties
   * Neighborhood patterns in stay duration requirements
   * Occupancy rates by stay strategy
3. **Pricing Factors**
   * Correlation between amenities and price
   * Impact of location on pricing
   * Relationship between accommodations and price

# Technical Implementation

### Data Processing Simplification

1. **Property Type Standardization**:
   * Reduce 50+ property types to 5 clear categories
   * Map all properties to standardized types
2. **Amenity Processing**:
   * Extract amenities from JSON arrays
   * Identify 15-20 most valuable/common amenities
   * Create binary feature matrix
3. **Stay Strategy Classification**:
   * Categorize properties as short or long-term focused
   * Calculate occupancy rates by strategy
   * Identify neighborhood patterns

### Focused Machine Learning

1. **Price Prediction Model**:
   * Use linear regression for interpretability
   * Include key property features and amenities
   * Evaluate with R² scoring
2. **Amenity Value Analysis**:
   * Extract feature importance from model
   * Rank amenities by price impact
   * Generate value-add estimates
3. **Stay Strategy Recommendation**:
   * Compare performance metrics by neighborhood
   * Develop decision logic based on property type
   * Create recommendation lookup function

### Streamlined Output

1. **Pricing Recommendation**:
   * Suggested price range based on property features
   * Property type categorization
2. **Amenity Suggestions**:
   * Top 3 high-value amenities to add
   * Estimated price impact for each
3. **Stay Strategy Guidance**:
   * Recommended minimum stay approach
   * Supporting neighborhood data

## Repository Structure

toronto\_airbnb\_pricing\_assistant/

├── data/

│ ├── raw/ # Original datasets

│ └── processed/ # Cleaned datasets

│

├── notebooks/

│ ├── 1\_data\_prep.ipynb # Data cleaning and feature engineering

│ ├── 2\_modeling.ipynb # Price prediction and feature importance

│ ├── 3\_analysis.ipynb # Stay length analysis and recommendations

│ └── 4\_demo.ipynb # User input and recommendation examples

│

├── src/

│ ├── data\_prep.py # Basic data processing functions

│ └── modeling.py # Model and prediction functions

│

├── results/ # Outputs and visualizations

│

├── individual\_work/ # Team member contributions

│ ├── sean/

│ ├── bryan/

│ └── jitesh/

│

├── requirements.txt

└── README.md

# Conclusion

The Toronto AirBnB Pricing Assistant & Profit Maximizer will provide AirBnB hosts and property investors with clear, actionable insights to guide their rental pricing and strategy 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 limited time availability and varying technical backgrounds.