**💡 Project Idea: Real Estate Market Analysis in Israel (Yad2 / Airbnb)**

**Overview**

The Israeli real estate market is known for its high prices, regional disparities, and strong demand. This project aims to analyze housing listings from **Israel** by scraping data from **Yad2.co.il**, the country’s leading platform for property listings. If access to Yad2 becomes limited or restricted due to anti-scraping measures, the project will pivot to **Airbnb**, focusing on short-term rental trends in Israeli cities.

**🎯 Goals**

* Collect real estate listing data from **Yad2 (preferred)** or **Airbnb (fallback)**
* Analyze price trends, common property types, and regional differences across **Israeli cities**
* Perform statistical and predictive modeling on housing prices
* Build an interactive dashboard to showcase key insights

**🛠️ Tools & Technologies**

* **Scraping**: Python, BeautifulSoup, Requests, Selenium
* **Cleaning & Analysis**: Pandas, NumPy
* **Database**: PostgreSQL, SQL
* **Modeling**: Statsmodels, Scikit-learn
* **Dashboard**: Tableau or Power BI

**🧩 Implementation Plan**

**Phase 1: Data Collection**

* **Primary Source**: Yad2 (Real Estate section - נדל"ן למכירה / להשכרה)
  + Extract: city, neighborhood, price (ILS), area (m²), rooms, floor, type, features (balcony, elevator, etc.)
* **Fallback Source**: Airbnb (if Yad2 access is blocked or limited)
  + Focus on listings in Israeli cities
  + Extract: location, price per night, number of reviews, room type, availability, rating, etc.

**Phase 2: Data Cleaning**

* Normalize Hebrew text
* Handle missing/duplicate entries
* Convert prices, area, room numbers to numeric
* Feature engineering (e.g., price per m², binary feature columns)

**Phase 3: Database Management**

* Create PostgreSQL schema for properties and locations
* Store clean dataset
* Write SQL queries to analyze average prices, property types, amenities

**Phase 4: Statistical Analysis**

* Descriptive stats and visualization
* Hypothesis testing (e.g., price difference between cities)
* Build regression models to predict price

**Phase 5: (Optional) Machine Learning**

* Train models to classify listings (e.g., affordable vs. expensive)
* Use Scikit-learn (Random Forest, Regression, KNN, etc.)

**Phase 6: Interactive Dashboard**

* Power BI or Tableau
* Visuals:
  + Map by city with average price
  + Filters for rooms, amenities, size, price range
  + Trends and comparisons

**Phase 7: Documentation**

* README / report with:
  + Project scope
  + Data source(s)
  + Cleaning process
  + Key findings
  + Screenshots and dashboard link
  + GitHub repo with code