

# **Build a database to streamline Dream Homes NYC's operations and elevate the client experience**

## **Final Project Presentation**

---

Yuxuan Chen (Iris)

Yishan Liu

Xianghui Meng

Haoxiang Huo(Steven)

[yc4637@columbia.edu](mailto:yc4637@columbia.edu)

[y15897@columbia.edu](mailto:y15897@columbia.edu)

[xm2346@columbia.edu](mailto:xm2346@columbia.edu)

[hh3109@columbia.edu](mailto:hh3109@columbia.edu)



# Introduction-

**“Our team will build Dream Homes NYC a centralized database that transforms unstructured records into our organized schema via an ETL process. The application on analytics and creation of Metabase dashboards enables faster, smarter decision-making, and gives agents and customers instant insights into listings, client behavior, and market trends.”**

## Task Division

Category	Task	Assigned Member(s)	Planned DDL
Business Understanding	Define detailed business requirements and client goals	Yuxuan Chen	Week 8
Database Schema	Design a normalized ERD (20+ tables in 3NF), including entities and relationships	Yishan Liu	Week 8
Data Simulation	Generate and clean sample data to simulate messy or semi-structured input	Steven Huo	Week 8
Table Creation	Write SQL DDL scripts to define all tables and constraints in PostgreSQL	Xianghui Meng	Week 8
Data Loading	Create Python/SQL scripts to load data into the database and write code for analytical procedure	Xianghui Meng	Week 9
Complex Queries	Write 8 complex SQL queries to support business decisions	Yuxuan Chen	Week 10
Dashboards	Design and build interactive dashboards using Metabase or Power BI	Steven Huo	Week 11
Customer Interaction Plan	Write Customer Interaction Plan	Yishan Liu	Weak 11
Final Report	Co-author project report, including intro, schema design, queries, and business insights	All Members	Week 12
Final Presentation	Prepare slides and live demo walkthrough; each member presents part of the work	All Members	Week 12

01.



# Business Insights:

1. Centralized Office Profiles
2. Comprehensive Employee Directory
3. Detailed Client Records
4. Full Property Listings
5. Appointment Management
6. End-to-End Transaction Tracking
7. Normalized, Relational Schema
8. Automated ETL Pipeline
9. Data-Driven Insights
10. Live Dashboards & Analytics

# Analytics Applications

## Data-Driven Real Estate Decision Solutions



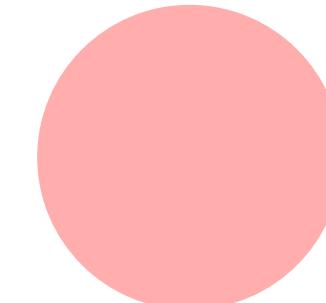
Dream Homes NYC transforms data into productivity through intelligent coding: Clients drag sliders on interactive heatmaps while backend executes `'price_range_analysis()'` to instantly recommend budget-matched listings, landlords input amenities to trigger `'rental_premium_calculator(gym, doorman)'` for precise rental yield forecasts. Marketing teams optimize ad spend using auto-generated ROI dashboards, while sales meetings dynamically allocate leads via real-time conversion funnels. Data evolves from static reports to an intelligent operational engine - powering every business process from client engagement to transaction closure.

# Conclusion

We transformed Dream Homes NYC's fragmented data into a scalable cloud-based warehouse with automated ETL, enabling real-time decision-making across all business functions while future-proofing for AI and IoT integrations.

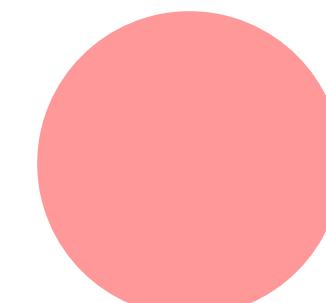
## Centralized Data Foundation

1. Consolidated scattered documents into a normalized relational database
2. Implemented foreign keys/constraints for 100% data integrity



## Self-Sustaining Automation

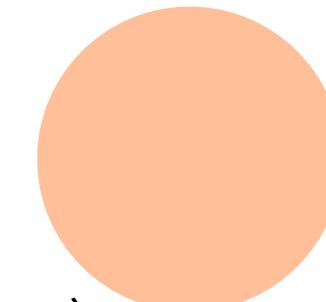
1. Python ETL pipelines with auto-rollback for nightly updates
2. Guaranteed report availability regardless of errors



## Role-Specific Intelligence

12 predefined queries answering:

1. Brokers' comps checks (2-sec mobile access)
2. Marketing's channel ROI tracking
3. Executive dashboards (revenue/ occupancy/ KPIs)



## Conclusion



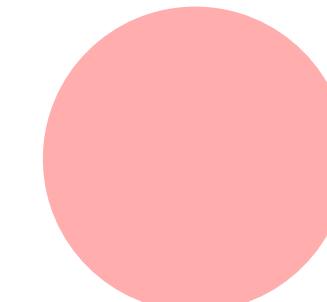
## Decision Velocity

1. Reduced analysis cycles from weeks → hours
2. Live transaction pricing during property viewings.



## Cloud-Native Scalability

1. Handles 10x query concurrency via horizontal scaling
2. Zero business code changes for infrastructure growth



## Future-Proof Architecture

- Ready for:
  1. Smart home IoT ingestion
  2. ML valuation models
  3. Self-service customer portals.

# ETL PROCESS SUMMARY

- EnhancedDreamHomesETL loads dream\_homes\_nyc\_dataset\_v8.csv into dream\_homes\_db on PostgreSQL.
- Semi-structured CSV records are converted to typed rows using stable natural keys for safe re-runs.

## Extraction:

- CSV read into pandas with UTF-8; empty cells → NaN; no line dropped.
- Original tokens preserved for typed converters in later stages.

## Cleaning & Validation:

- Safe type converters (safe\_int, safe\_decimal, etc.) return NULL and log warnings on failure.
- Missing monetary fields skip Transaction insert but still load related Property.

## Transformation:

- Normalize composite fields (e.g., addresses, bed/bath, client blurbs).
- Map categories via enum mappers to prevent spelling drift.
- Expand lists into child rows (e.g., features, appointments, documents).
- Fallback logic: missing offer\_date → use listing\_date or today; similar for offer\_amount.





## Database Design Summary

Schema follows third normal form (3NF) with enforced referential integrity.

Key relationships:

- Office (1) → Employee (M)
- Client (1) → ClientRole (M)
- Property (1) → Feature (M), Media (M)
- Transaction (1) → Commission (1) → Lease (1) → PaymentRecord (M)
- Appointment connects agent, client, and property in time-stamped interactions.
- Documents and MarketingCampaigns link to properties or transactions.
- Natural keys (e.g., email, mls\_number, transaction\_code) support safe, deduplicated upserts.
- Design supports complex workflows across both sales and rentals











Thanks for watching!