



Housing Price & Salary

Shum Chang Wu Chen (Willy)
Yucheng Chen
Alex Moronta
Jubayeth Mian

Baruch COLLEGE | **CUNY**



Goal

- To determine if properties are within budget range based on average salaries from various professions within the NY county.

Importance

- This will allow users to identify the ideal location to look for properties based on expected salary within their profession.



Business Requirements

- The most important factors for our users are:
Salary & Housing Price Range
- Dashboard showing measures that fall into the desired price range, location, etc.



Functional Requirements

- **Business Intelligence Usage**
Generate tables with relevant information for report & and graphs for trends/pattern inferences.
- **Detailed Analytics**
Generating refined data columns for statistical purposes such as regression models & correlation analysis.



Data Requirements

- **Geographical Coverage**

Data on properties will cover those in NY, NJ, CT and PA.

- **Processing**

String values for keys must be as consistent and standardized as possible across all tables. Null values in required fields should be filtered out.

- **Pipeline**

Data should be stored in one central warehouse (Snowflake).



Regulatory Requirements

- **User Privacy**

Seller and buyer information should not be shown or used.

- **Data Sourcing**

All information must only be sourced from publicly accessible locations to avoid confidential business info from being leaked.



Business Personas

- **Homebuyers**
Looking to for relocation options near New York or first time house.
- **Real Estate Agents**
Information gathering for their operating region/area.
- **Data Scientists**
To determine variable impacts on price



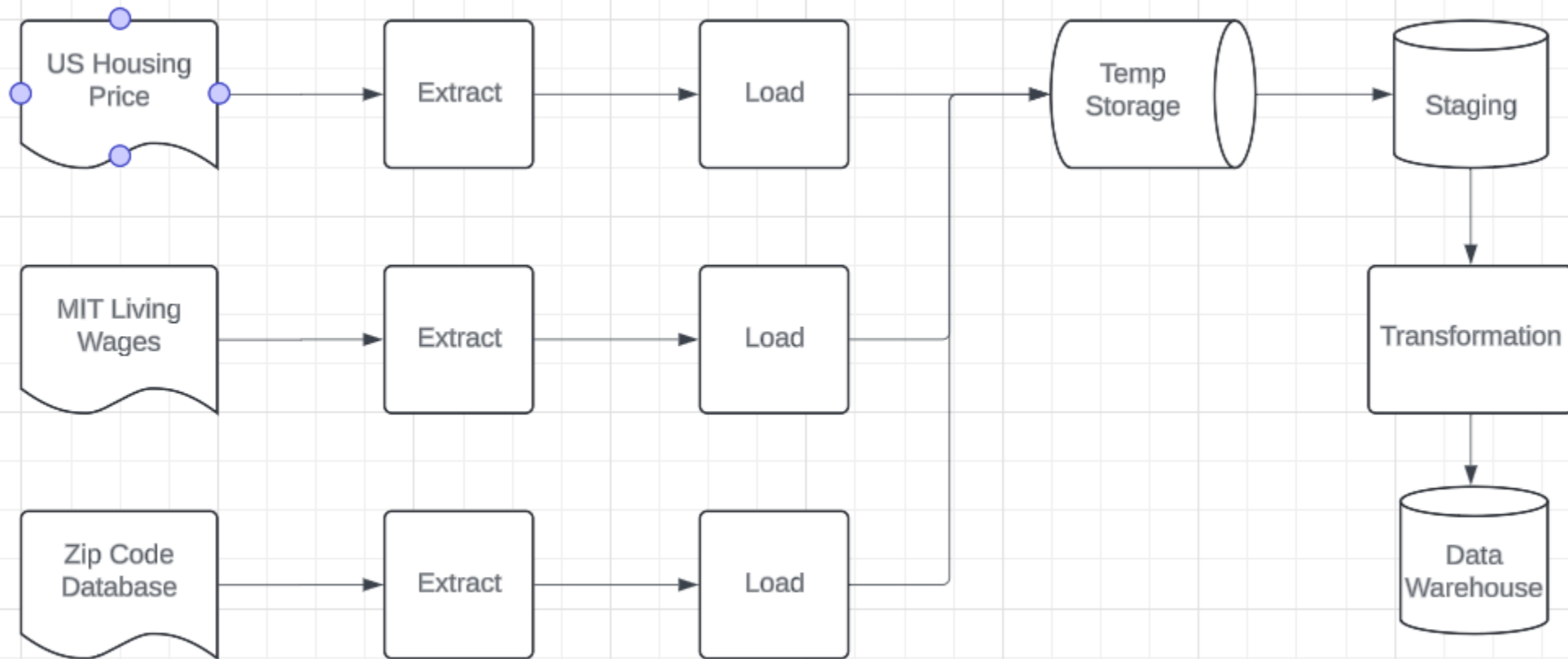
Benefits

- Identifying pricing trends over the a time period.
- Monetization of database for purpose of developing real estate apps & services.
- Allows the integration of future housing data in other regions.

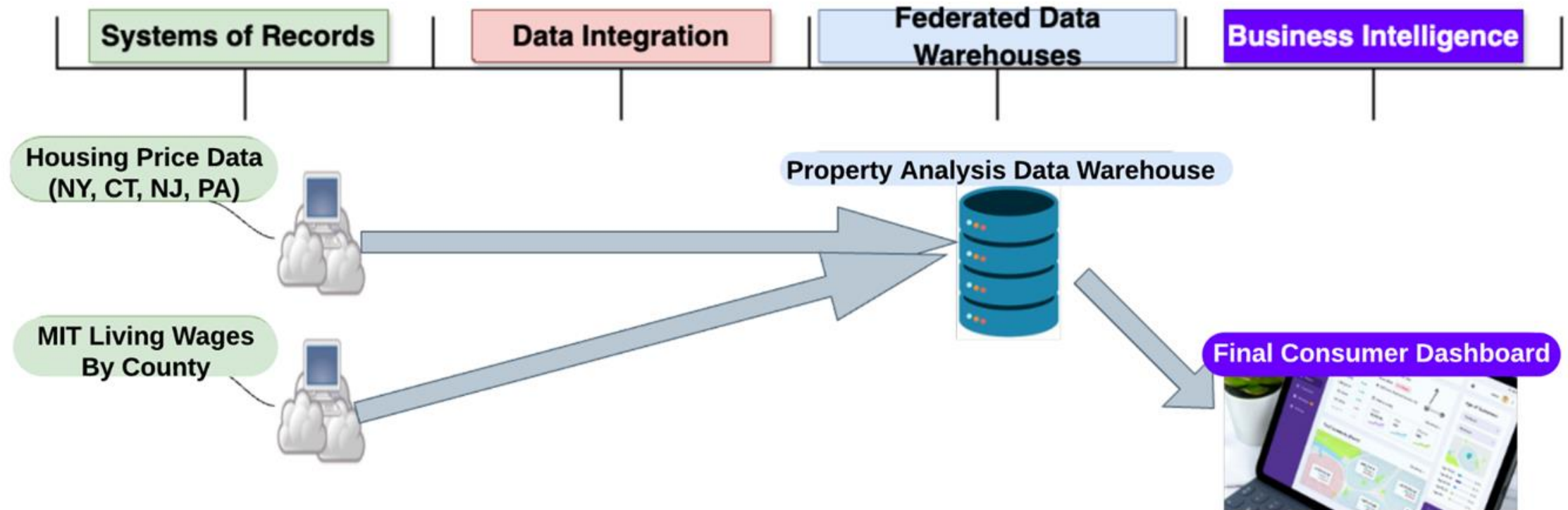
Costs

- Server upkeep and usage costs
- Time investment to develop and create the data pipeline and its BI uses.
- Opportunity cost in resources spent

Information Architecture

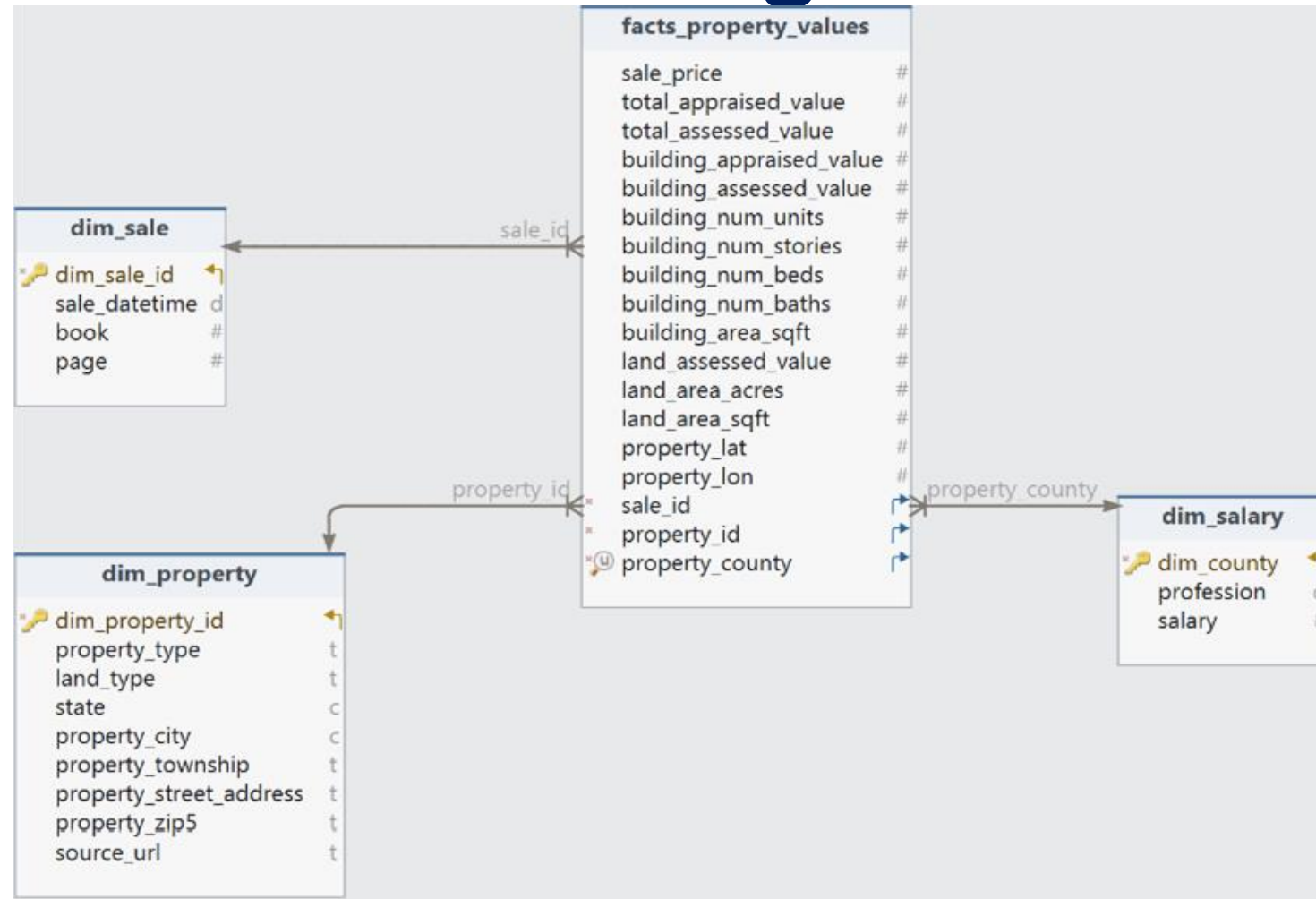


Data Architecture

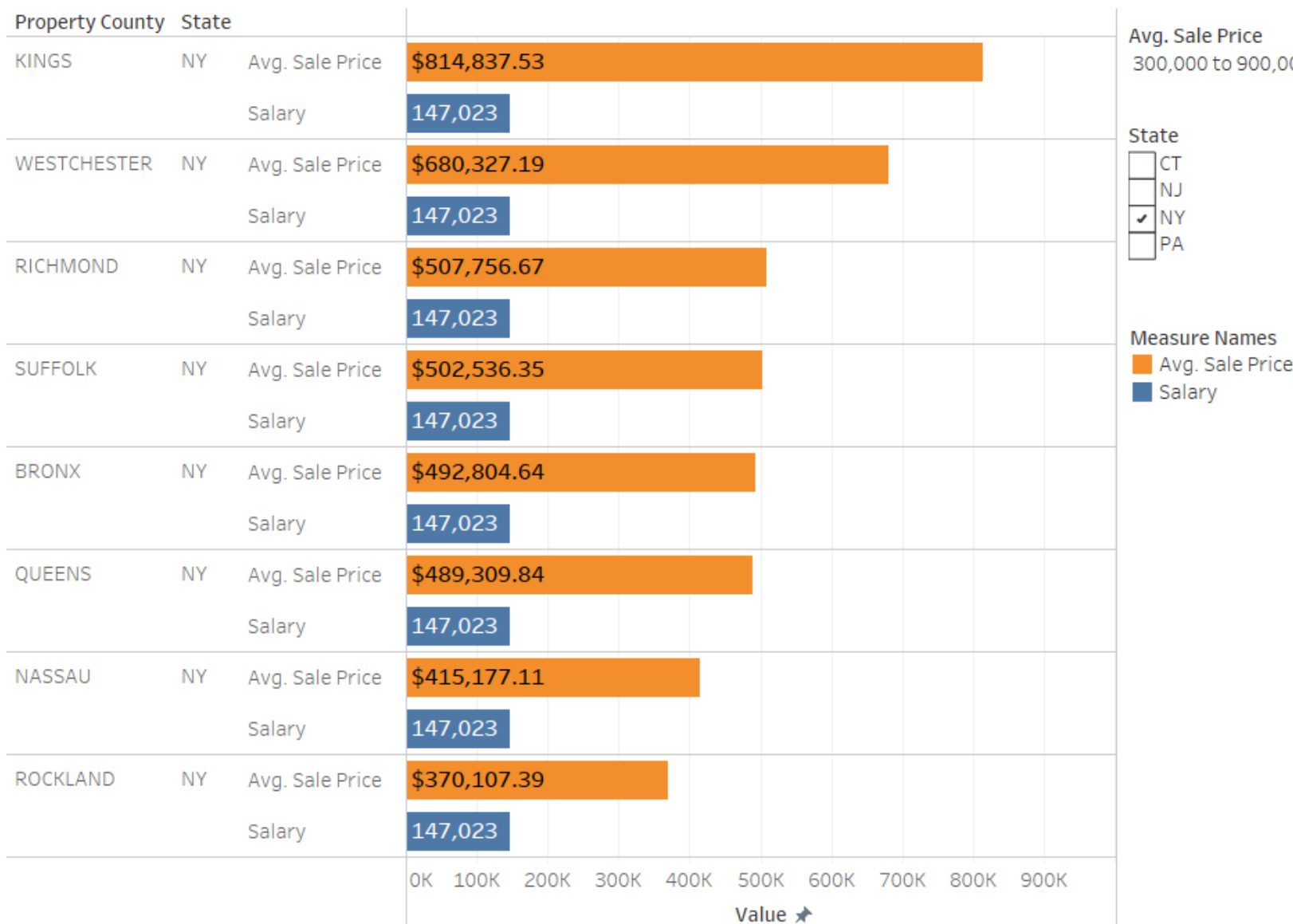




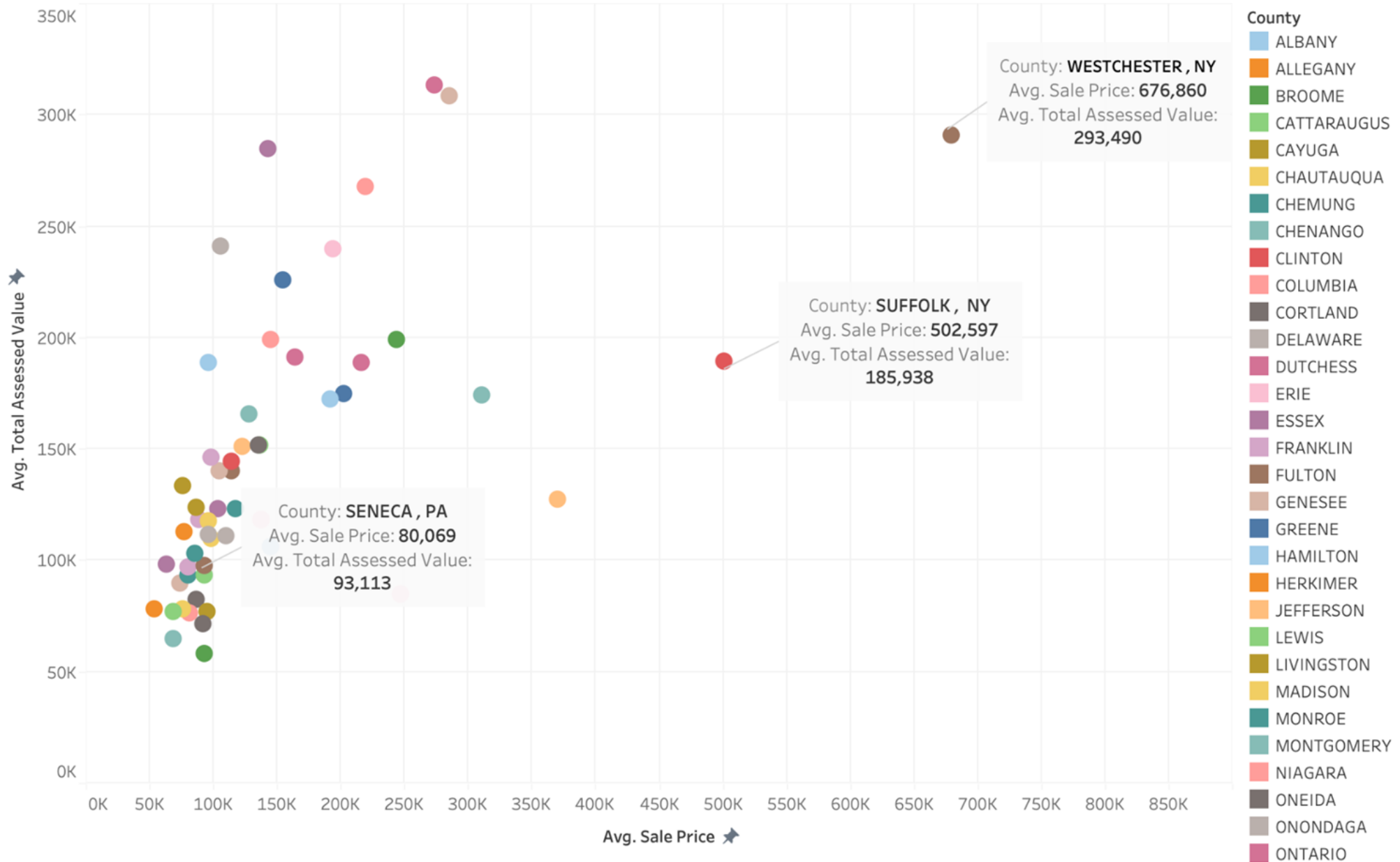
Dimensional Modeling



Property By Avg Price & Profession Avg Salary



Sale Price Compared to Assessed Value





Conclusion

- Houses in the New York City area and its neighboring counties are extremely expensive.
- Affordability by professions:
Management, Computer Mathematics, Business & Financial Operations, Legal, Healthcare Practitioners, Technical & Architecture & Engineering



Future Work

- Expand our data warehouse with the following data:
 - States outside of the four we covered
 - Cost of living data.
 - More comprehensive salary data and perhaps mortgage data.
 - Housing metrics such as number of bedrooms, bathrooms, backyard, etc.



Members Contribution

Shum Chang Wu Chen (Willy)

- Information Architecture
- Data extracting (NY)
- Azure & Snowflake

Yucheng Chen

- Data extracting (NJ)
- Dimensional Modeling
- Snowflake & DBT

Alex Moronta

- Data Architecture
- Data extracting (PA)
- Data Visualization

Jubayeth Mian

- Data extracting (PA)
- Presentation

A dark blue, monochromatic image of a city skyline, likely New York City, with various skyscrapers and buildings visible against a slightly lighter blue sky. The image has a grainy, textured appearance.

Thank You!

I hope you can find your dream home one day.

