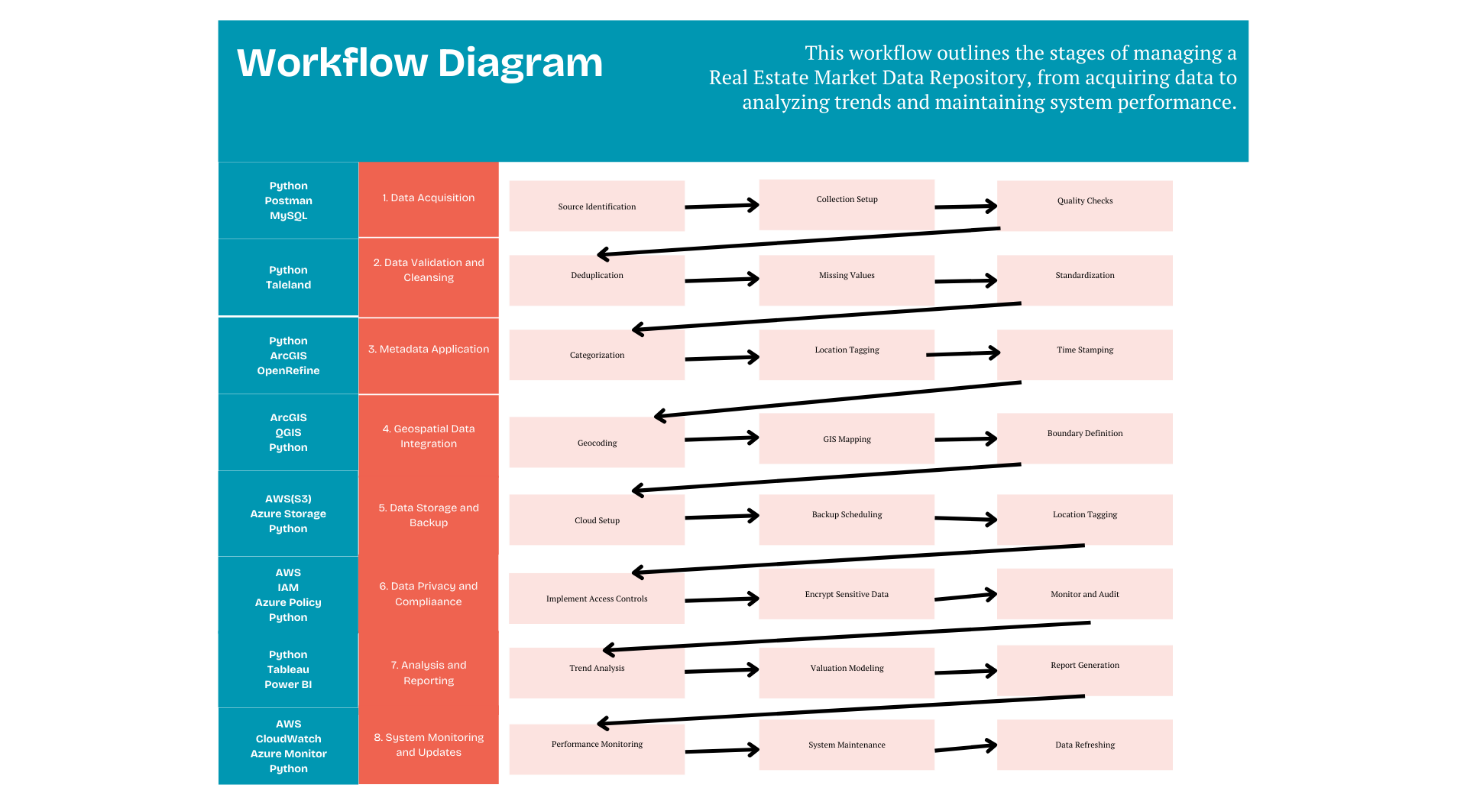
**Nashville Housing Data Insights and Predictive Analytics Repository**

**Name:** Atajan Arashev  
**Date:** December 2024  
**Repository URL:** [MySQL Real Estate Analysis](https://github.com/AArashev/MySQL-RealEstate-Analysis)

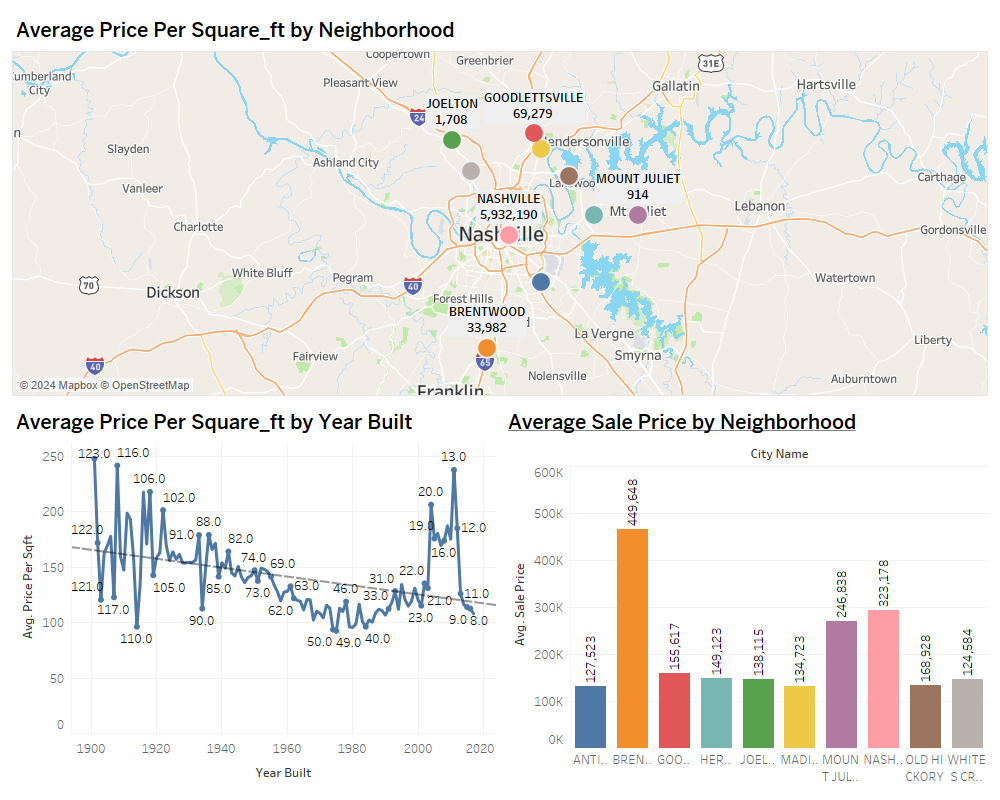
**Introduction**



Digital curation and preservation ensure the usability, authenticity, and accessibility of digital data over time. In the real estate industry, decision-making relies on the availability of accurate and structured data to analyze trends, forecast prices, and support investments.

The **Nashville Housing Data Insights and Predictive Analytics Repository** is designed to curate and preserve property data, including transaction records, valuations, and predictive insights. By integrating machine learning models, geospatial tools, and interactive dashboards, the repository supports decision-making for real estate professionals, investors, and researchers.

**Items or Data Preserved/Curated**

****

**Type of Information:**

* **Content**:
  + Property sales records, including sale prices and property details.
  + Predictive insights using machine learning models (e.g., XGBoost and ANN).
  + Geospatial data for neighborhood analysis.
* **Format**:
  + SQL tables for structured data storage.
  + Tableau dashboards for visualizing trends and insights.

**Preservation Plan:**

* Data is stored in a cloud-based MySQL database with automated backups and regular updates.
* Machine learning predictions are periodically retrained and integrated into the database.
* Geospatial data is linked with property records for neighborhood-level analysis.

**Potential Users**

**Target Audience:**

* **Real Estate Professionals**: For analyzing market trends and pricing strategies.
* **Investors**: For identifying undervalued properties and high-growth areas.
* **Researchers**: For studying property appreciation and urban development trends.

**Access and Use:**

* Users can retrieve data through pre-defined SQL queries.
* Tableau dashboards provide visual insights into market trends and predictive analysis.
* Future development will include API integration for seamless real-time access.

**Software/Tools Used**

* **MySQL**: For storing and querying structured data.
* **Python**: For data preprocessing, machine learning, and database integration.
* **Tableau**: For creating interactive dashboards that visualize market trends.
* **ArcGIS**: For geospatial data integration and analysis.
* **GitHub**: For version control and sharing the repository.

**Metadata Fields**

**Fields and Descriptions:**

1. **property\_id**: Unique identifier for properties.
2. **sale\_price**: Recorded sale price of the property.
3. **valuation\_to\_sale\_price\_ratio**: Ratio comparing valuation to sale price.
4. **city\_id**: Identifier for the property's city.
5. **predicted\_sale\_price**: Sale price predicted by machine learning models.
6. **price\_per\_sqft**: Sale price divided by the property’s square footage.

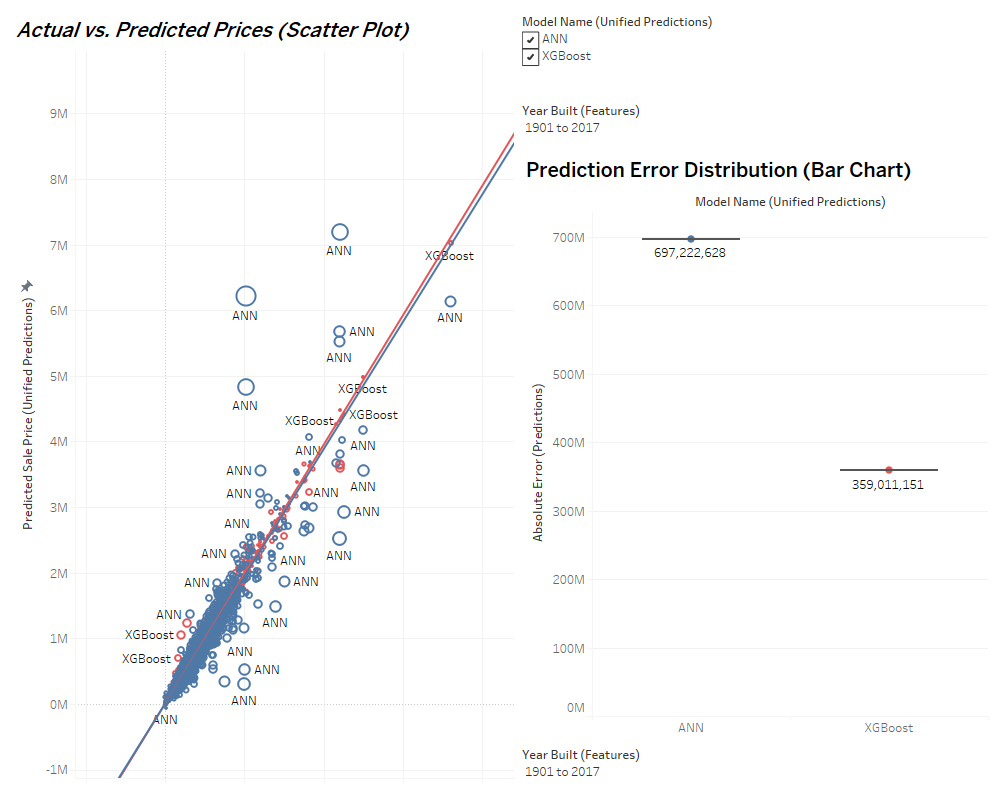
**Copyright**

* **Ownership**: Data curated from public sources or synthesized for educational purposes.
* **User Responsibility**: Users must cite the repository when using its contents in publications or reports.

**Workflow and Maintenance**

**Steps to Create the Repository:**

1. Data collected from public datasets and imported into MySQL using Python.
2. Database schema designed using the EER diagram for normalization.
3. Machine learning models (XGBoost and ANN) trained, and predictions stored in the database.
4. Tableau dashboards were created for visualizing trends and actionable insights.

****

**Maintenance Schedule**

* **Quarterly**: Update datasets and retrain predictive models.
* **Annual**: Conduct metadata reviews and update fields based on user needs.

**Budget:**

|  |  |  |
| --- | --- | --- |
| Category | Expense | Yearly Cost |
| Personnel | Data Analysts, GIS Specialists, etc. | **$398,850** |
| Software and Tools | Database licenses, Tableau, ArcGIS | **$10,000** |
| Infrastructure | Cloud storage, development, API hosting | **$12,000** |
| Miscellaneous | Data acquisition, training, marketing | **$7,200** |
| Total |  | **$428,050** |

**Conclusion**

**Reflections:**

The repository faced challenges such as selecting suitable machine learning models and integrating predictions into a normalized database. These challenges were resolved by implementing automated pipelines and leveraging visual tools like Tableau for clear insights.

**Future Enhancements:**

1. Develop APIs to enable real-time data retrieval and integration.
2. Expand dashboards with geospatial clustering for neighborhood-level insights.
3. Incorporate ensemble learning models for enhanced prediction accuracy.

**References**

1. Arashev, A. (2024). **MySQL Real Estate Analysis Repository**. GitHub. Retrieved from <https://github.com/AArashev/MySQL-RealEstate-Analysis>
2. Stratify Workflows. (2024, March 20). **How to use workflows for collaborative budgeting**. Stratify Blog. Retrieved from <https://www.stratifytech.com/blog/how-to-use-workflows-for-collaborative-budgeting>

**Appendix**

**Metadata Example:**

|  |
| --- |
| **property\_id**: 1001 |
| **sale\_price**: $350,000 |
| **valuation\_to\_sale\_price\_ratio**: 0.9 |
| **city\_id**: 1 |
| **predicted\_sale\_price**: $360,000 |
| **price\_per\_sqft**: $125.50 |

