**Table of Contents**

[Earn more money through real estate investments 1](#_sjtzv2elbefz)

[Table of Contents 2](#_ct2nip7pr7p8)

[1. Business Objective 3](#_hn7mkbkdvej6)

[2. Key actionable business initiatives 3](#_o3evl1mk93ub)

[3. Metrics of success 4](#_n8hg7lwzjcsb)

[4. Role of analytics 4](#_jlvr964qtb1n)

[5. Thinking through the analytics 5](#_t0p2e0lcu3kz)

[6. Executing the Analytics 7](#_425t0ltnyrjl)

[7. Implementation 7](#_vclao2pp0tac)

[8. Scale 9](#_kdocsfnif2x)

[9. Use our model to strategize 10](#_pk1fwzxawijm)

# **1. Business Objective**

Our main business priority is to transform the rental and real estate industry with the help of natural language processing for better investments and earning more profits. We aim to achieve:

* First, to develop a predictive model that accurately forecasts house prices
* Second, to offer prescriptive advice to maximize property values and aid buyers in making informed decisions in the Austin housing market

# **2. Key actionable business initiatives**

1. **Unified Data Management:**
   * We can streamline data collection from various real estate sources like Zillow listings, property databases, and economic indicators into a centralized data lake on the Google Cloud Platform (GCP)
2. **Prospective Insights Dashboard:**
   * Developing an interactive dashboard on GCP using Data Studio or Looker that provides real-time insights into housing market trends, price forecasts, and investment opportunities
3. **Automated Models:**
   * Implementing machine learning algorithms on GCP's AI Platform to build models that estimate property values based on historical sales data, location factors, and market trends.
   * Enabling real estate professionals to quickly assess property valuations, perform risk analysis, and optimize pricing strategies for better ROI.
4. **Scalable Model Deployment:**
   * Deploying scalable APIs on Google Kubernetes Engine (GKE) to handle high volumes of valuation requests from real estate agents, investors, and financial institutions.
   * Ensuring low-latency responses and high availability of valuation services, which streamlines property transactions and accelerates investment decision-making processes.
5. **Performance Monitoring and Optimization:**
   * Implementing monitoring dashboards on GCP's Cloud Monitoring to track model performance metrics, detect anomalies, and optimize model accuracy and reliability over time

The most impactful business initiative here would be Scalable Model Deployment as we can use it for large datasets and ever-growing amounts of information. We can use it for other cities and impose it company-wide to enable better business growth. It would act as a trusted advisor by our side, giving us the right information at the right moment to grow our business and increase profitability in the real estate sector.

# **3. Metrics of success**

Metrics in the order of their priority are:

1. **Prediction Accuracy**: This remains a primary metric, focusing on the precision of the predictive model in estimating current and future house prices. Accuracy can be measured using metrics like RMSE (Root Mean Squared Error) and MAE (Mean Absolute Error).
2. **Feature Importance Validity**: Measure how accurately the identified features predict price changes over time. This can be evaluated by back-testing: using historical data to see if the features previously identified as important accurately predicted the trends and prices.
3. **Economic Impact**: Specifically for investors, measure the return on investment (ROI) from properties bought based on model predictions. For real estate agents, track the sales volume or the speed of sales transactions as metrics of success.

**Hypothesis**: If the model accurately identifies and ranks the most relevant features affecting housing prices, then investments made based on these insights will yield a higher ROI compared to the market average, and real estate agents will experience quicker sales cycles and higher transaction volumes.

# **4. Role of analytics**

Analytics might add value

1. **Enabling the Business Initiative:** Analytics forms the backbone of our business strategy, providing the technological and methodological foundation necessary for executing our key initiatives:

* **Predictive Modeling:** By utilizing advanced machine learning algorithms, analytics enables the creation of models that accurately predict house prices based on a variety of factors. This predictive capability is essential for guiding investment decisions and advising clients.
* **Prescriptive Analysis:** Analytics also supports the development of tools that offer prescriptive advice to property owners and buyers. This helps in maximizing property values and ensuring that buyers make informed decisions based on robust data analysis.

1. **Refining the Business Initiative:** The iterative nature of analytics means it plays a crucial role in continuously refining our business strategies:

* **Feature Refinement:** Ongoing analysis helps identify which features of properties most significantly affect their market value, allowing for the continuous refinement of our predictive models.
* **Market Adaptation:** Analytics enables us to quickly adapt to changes in the real estate market by providing insights into shifting trends and consumer behaviors. This agility is critical for maintaining competitiveness and relevance in a dynamic market.
* **Feedback Utilization:** Incorporating feedback from both the analytics tools’ users and market performance data, analytics drives the refinement of business initiatives, ensuring they are aligned with actual market needs and user preferences.

1. **Evaluating the Success of the Business Initiative:** Analytics is integral to evaluating the effectiveness and efficiency of our initiatives, providing a clear measure of success and areas for improvement:

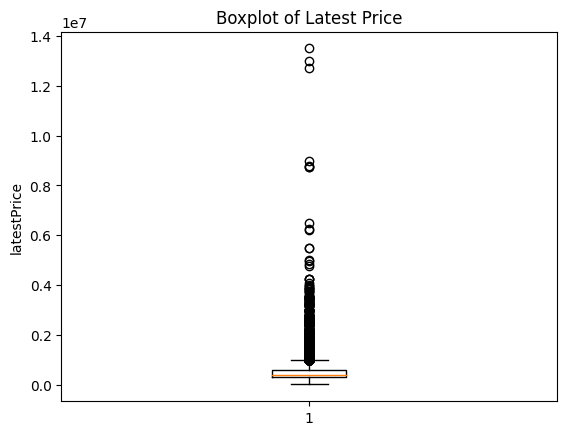
* **Performance Metrics:** Through analytics, we establish and monitor key performance metrics such as model accuracy, user engagement, and economic impact. These metrics are vital for assessing the success of our initiatives and guiding future strategies.
* **Outcome Analysis:** Analytics helps quantify the business outcomes of our initiatives, such as the ROI from properties selected based on our predictive insights and the effectiveness of our prescriptive advice in enhancing property values.
* **Strategic Adjustments:** By analyzing the outcomes and ongoing performance metrics, analytics provides the basis for making informed strategic adjustments, ensuring that our initiatives continue to deliver maximum value to our clients and stakeholders.

# 

# **5. Thinking through the analytics**

1. **Data**

* **Data Source:** Our data is derived from an open data source (Kaggle). The dataset is a comprehensive record of Austin Texas housing sales from 2018 to 2020.
* **Outcome and explanatory variables**: The outcome variable we are interested in is the ‘latestPrice’, which records the latest deal price of a house sold on a specific date. As showcased below, the outcome variable is skewed to the right on a slight scale, meaning that many houses are sold at extremely high prices but none are sold at extremely low prices, using outliers defined by IQR as the standard.



There are plenty of explanatory variables, some are categorical (e.g., home type), some are numerical (e.g. living area), and some are boolean (e.g. hasCooling). We labeled the categorical variables and combined them with numerical variables to form the combined training features.

1. **Type of analytics**

We mainly found that the exploratory analytics and predictive analytics add value to our analysis. EDA provides a deep understanding of the distribution of our data, uncovering key patterns, relationships, and outliers that inform subsequent analysis. It aids in feature selection and transformation, ensuring the most relevant and clean data is used for predictive modeling. Predictive analytics generates accurate forecasts of housing prices, aiding in decision-making for investments and pricing strategies. It identifies key drivers of housing prices and enables scenario analysis, enhancing strategic planning and operational efficiency. Our analysis is not very causal-oriented since there exists severe multicollinearity in our features.

1. **Predictive analytics**

If your analytics is predictive, do you care more about the best prediction or about

answering "why" questions (or both)? Based on this answer, do you want to focus on

analyst or data-driven models (or do you need both)?

Both the best prediction and answering “why” questions are important for us. In the analysis, we tried to find the R-squared value for the XGBoost model that we built to evaluate how good our model is in terms of the variation explained. We also analyzed the feature importance which explains what are the top features that drive the change in price.

1. **Impediments**

Feature Engineering Challenges: Creating meaningful features from raw data is complex and time-consuming. We conducted a thorough EDA to identify potential new features. Used domain knowledge to create additional relevant features by combining similar features (e.g. number of different types of schools). Encoded some features to incorporate categorical features into our model.

# **6. Executing the Analytics**

• Who in your organization is going to be responsible for executing the analytics? • Who will collect the data? Who will run the models? Who will implement and evaluate the experiments? • How will you involve them in defining metrics and thinking through the analytics?

Our company will have a dedicated Data Analytics Team comprising data scientists, data analysts, and data engineers who are responsible for executing the analytics for the Austin house price prediction project.

* **Data Collection:** Data engineers will be responsible for collecting, cleaning, and maintaining the data. They will ensure that the data is accurate, up-to-date, and ready for analysis.
* **Model Development and Execution:** Data scientists will take the lead in developing predictive models. They will use statistical and machine learning techniques to build and refine models based on the project requirements.
* **Experiment Implementation and Evaluation:** Data analysts will assist in running the models and implementing the experiments. They will also be responsible for evaluating the outcomes of these experiments, comparing them against the set metrics, and providing insights.
* **Defining Metrics and Analytics Thinking:** The involvement of the entire Data Analytics Team in defining metrics and refining the analytics approach is crucial. Regular meetings will be organized where team members can present their findings, discuss discrepancies, and make collective decisions on the direction of the analytics initiatives. Stakeholder feedback will also be integrated into these discussions to ensure that the analytics are aligned with business objectives and the real needs of the market.

# **7. Implementation**

1. Once you have the results, what decisions will this influence? – What would do differently because of the analytics developed?

**Price strategy:**

* With insights into the most important features influencing property prices, such as living area size and price per square foot, pricing strategies can be more accurately formulated. This ensures properties are neither overpriced nor underpriced.
* Develop more sophisticated pricing models that take into account the most influential features such as living area size and price per square foot. This could involve dynamic pricing adjustments based on real-time market data and property characteristics.

**Marketing and Sales:**

* Create marketing campaigns that highlight the features identified as most valuable by buyers. For example, emphasizing properties with larger living areas or those located in high-demand areas like Austin, TX.
* Personalize property listings by prominently featuring the most important characteristics, such as living area size, number of bathrooms, and price per square foot. This can make listings more appealing and relevant to potential buyers.

**Property Investment:**

* Make informed decisions on property investments by prioritizing properties that have the key features identified as most valuable (e.g., larger living areas, desirable locations).

1. Have you thought through about existing or new workflows and how would you embed analytics to ensure adoption?

**Analysis Current Processes:**

* Conduct a thorough review of existing workflows in areas such as pricing, marketing, sales, and investment decision-making. Identify where and how decisions are currently made and what data sources are used.
* Pinpoint specific stages within these workflows where analytics can provide valuable insights. This might include price setting for new listings, marketing campaign planning, investment evaluations, and customer interaction.

**Develop New Workflows:**

* Ensure that data collection, analysis, and insights generation are integral parts of these processes. For example, a new workflow for property pricing might start with data extraction, followed by model application, and conclude with an automated pricing recommendation.
* Implement pilot programs for these new workflows in select regions or departments. Gather feedback, refine the processes, and ensure that the analytics provide actionable insights that improve decision-making.

**Ensure Adoption:**

* Embed analytics directly into the tools and platforms that employees use daily, such as CRM systems, property management software, and sales platforms. This ensures that the insights are readily available and easy to use.
* Conduct comprehensive training sessions for all users on how to interpret and utilize the analytics. Provide ongoing support and resources, such as user manuals, FAQs, and help desks, to assist with any issues or questions that arise

**Embed Analytics:**

* Sales and Marketing: Embed analytics in CRM systems to provide sales teams with real-time insights on property values and customer preferences.
* Pricing Strategy: Integrate pricing analytics into the property listing process. Ensure that every new listing is priced using the developed models, and adjustments are made automatically based on market trends.
* Investment Analysis: Embed analytics into investment evaluation tools, providing investors with a clear view of the potential return on investment based on the key property features identified.
* Customer Interactions: Use analytics to personalize customer interactions by recommending properties that match their preferences and budget, enhancing the customer experience and increasing the likelihood of successful transactions.

# **8. Scale**

**Challenges in Scaling Real Estate Analytics Initiative:**

1. **Data:**
   * Inconsistent data from multiple sources.
   * Ensuring data accuracy and privacy.
2. **People:**
   * Skill gaps in analytics and real estate modeling.
   * Resistance to adopting new tools.
3. **Systems:**
   * Legacy IT infrastructure and scalability issues.
4. **Culture:**
   * Reliance on traditional methods.
   * Shifting to data-driven decision-making.
5. **Budget:**
   * Scaling the analysis, for example in a cloud environment, can tremendous fee that exceeds the budget allocated to the analytics.

**Addressing Challenges:**

1. **Data:**
   * Implement data governance policies.
   * Use ETL tools for data integration.
   * Enhance data security measures.
2. **People:**
   * Offer training programs in analytics.
   * Engage stakeholders early to gain buy-in.
3. **Systems:**
   * Upgrade existing data to scalable, cloud-based infrastructure.
   * Gradually integrate new tools with existing systems.
4. **Culture:**
   * Use change management strategies.
   * Highlight successful data-driven case studies.

**Continuous Improvement Plan:**

* Adopt an iterative model development approach.
* Establish a dedicated analytics team.
* Stay updated with advancements in real estate analytics.
* Implement KPIs to measure and improve model impact.

# **9. Use our model to strategize**

Our predictive model using XGBoost provides accurate housing price predictions, allowing stakeholders to make informed decisions. Real estate agents can use these predictions to set competitive prices, ensuring faster sales and higher client satisfaction. Investors can identify undervalued properties by comparing predicted prices with market listings, optimizing their investment portfolios. We can increase our prediction timeframe from currency 1 year to more as needed. The TF-IDF text analytics reveal the most frequent words in property descriptions, highlighting key features that attract buyers. Marketers can leverage this insight to craft compelling listings emphasizing these attributes, enhancing buyer interest. By combining price predictions with text analytics, stakeholders can strategically target improvements and marketing efforts, maximizing returns and efficiency in the Austin housing market.

**Appendix:**

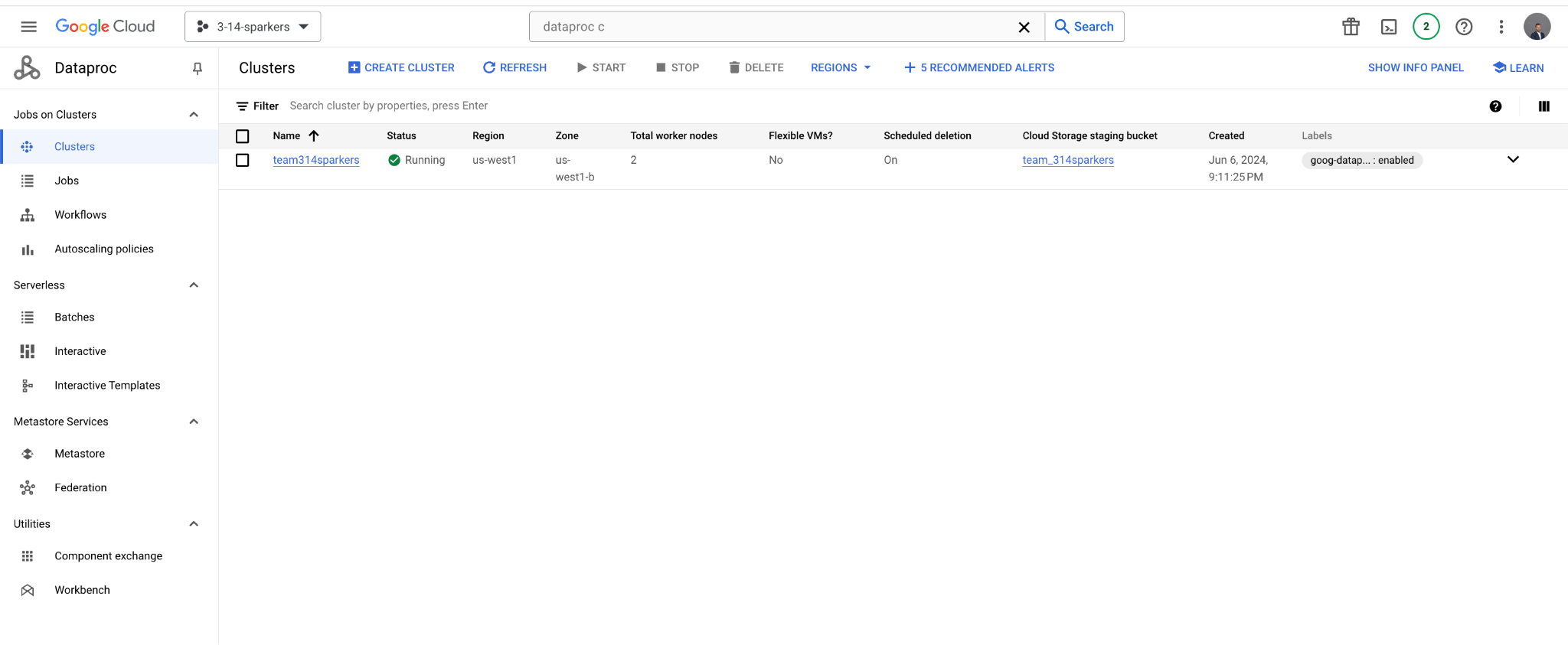
**Appendix:**

**Google Cloud Platform:**

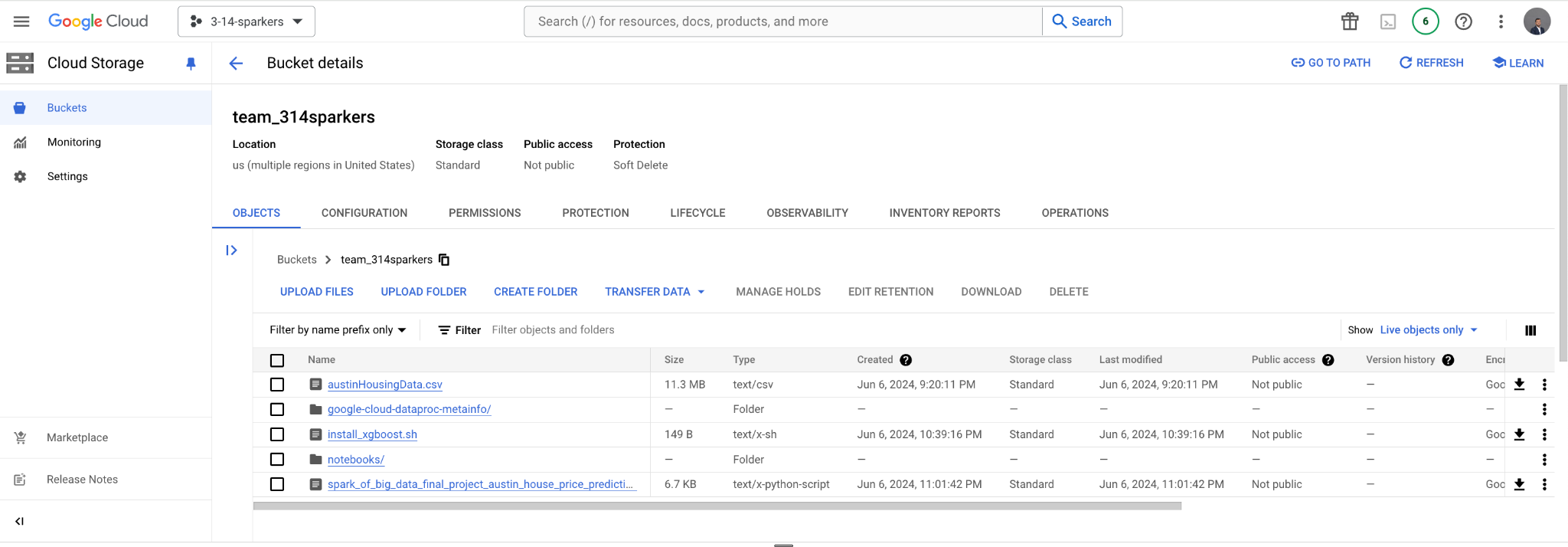
Please find below screenshots of the Google Cloud Platform steps we used to run our pyspark job on the cloud. This method helps us scale the data and ensure automation by creating jobs and pipelines

**Steps of using Google Cloud Platform for creating clusters and spark job:**

**Creating clusters:**

****

**Uploading files to clusters:**



**Running Spark Job:**



**Spark job output:**