|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | | | |
|  |  |  | |  |
|  | Affordable Housing Market Analysis for Strategic Expansion  strategic analysis report | | |  |
|  | fictional housing solutions inc. | | Prepared by kALLE rEAVES |  |

|  |  |  |  |
| --- | --- | --- | --- |
|  | | | |
|  |  |  |  | |
|  | executive summaryproject Purpose There has been a lot of talk about unaffordable housing is for this generation of Americans.  It’s no secret that we have felt the effects of this as a company. We have had numerous layoffs, a drop in revenue growth. Our model of selling houses has been overshadowed by larger corporations who do are not even in the business of selling houses buying up the inventory and driving up prices. The purpose of this project is to find an affordable solution to these issues for both Financial Housing Solutions and the customer. Summary of findings From Just from 1992 housing prices have increased 350% while hourly wages have increased only increased about 100% during the same time. The average houses price in 2025 is now 416900$ while the average hourly wage excluding 1% earners is around 20$/hr. The simplest calculation to use to decide if someone can afford a home is (House price/4.5 <= annual wage) When the average annual wage is around 42000$ a year that means most people would have to earn twice as much! This doesn’t include everyday expenses like paying their rent, electricity bill, car note, food etc. This of course is bad for our business since we are in the business of selling these houses Key recommendations  * We need to lower the prices of houses that are 3 bedrooms or less. This will bring in more revenue faster. * Second, we must take advantage of lower cost areas to build and sell houses.  Tools Used Python language, Jupyter Notebook, VScode, Dbeaver community, Mssql, Apache Superset, Microsoft word, Notebook ++, Ubuntu 24.04, Windows 11 pro  Python libraries: Sklearn, Matplotlib, Sklearn, sqlalchemy, Pandas, Numpy | |  | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | | |  | | | |
| Business Case  The company has invested heavily in research and development, with a focus on innovation and sustainability. This year, Fictional allocated $50 million towards housing initiatives to improve sales objectives Fictional Housing solutions seeks to create a stable affordable market for housing by improving business strategy and making data driven insights Assumptions/constraints Data was readily available from many sights including but not limited to Kaggle, Bureau of labor statistics, and Zillow.  Schedule:  - Project Start: May 29, 2025 - Mid-point Check-in: June 15, 2025 - Project End: August 1, 2025 - Key Milestones completed:  • Data Collection (Week 3)  • SQL Schema Design & Population (Week 4)  • EDA & Cleaning (Week 5)  • Feature Engineering (In Progress - Week 6)  • Modeling (Starting Week 6)   | **Stakeholders** |  | | --- | --- |      |  |  |  | | --- | --- | --- | | |  |  | | --- | --- | | **Project Sponsor** | Kalle Reaves | | | |  |  | | --- | --- | | **Project Manager** | Kalle Reaves | | | |  |  | | --- | --- | | **Data Analyst / Scientist** | Kalle Reaves | | | |  |  | | --- | --- | | **BI Developer** | Kalle Reaves | | | |  |  | | --- | --- | | **Database Administrator (DBA)** | Kalle Reaves | | | |  |  | | --- | --- | | **Software Engineer** (optional) | Kalle Reaves | | | | | | | |
|  |  |  | |  |  |

|  |  |
| --- | --- |
|  |  |
| work breakdown structure **1. Project Planning & Management**  1.1 Define project objectives and scope 1.2 Identify data sources (Kaggle, Zillow, Census, BEA, etc.) 1.3 Develop project timeline 1.6 Schedule regular progress checkpoints  **2. Data Collection**  2.1 Research housing price datasets 2.2 Download Zillow historical housing price data 2.3 Collect inflation rate data from BEA 2.4 Gather interest rate data from Federal Reserve 2.5 Collect demographic/economic data from U.S. Census 2.6 Acquire property-level data with location and features 2.7 Document metadata and data sources  **3. Data Storage & Organization**  3.1 Design SQL database schema 3.2 Create SQL tables for each dataset 3.3 Populate tables with collected data 3.4 Normalize and index SQL tables 3.5 Back up SQL database  **4. Data Cleaning & Preprocessing**  4.1 Handle missing values 4.2 Remove duplicates 4.3 Standardize data formats  4.4 Convert SQL tables to pandas DataFrames 4.5 Perform exploratory data analysis (EDA) 4.6 Merge data across sources by ZIP code/state  **5. Feature Engineering**  5.1 Create new variables (e.g., price per square foot) 5.2 Encode categorical variables 5.3 Create dummy variables for amenities 5.4 Normalize/standardize numerical features 5.5 Generate time-based features for time series analysis  **6. Statistical & Predictive Analysis**  6.1 Conduct correlation analysis 6.2 Perform linear regression to assess price factors 6.3 Run multivariate regression analysis 6.4 Conduct time series analysis  6.5 Test models for inflation-adjusted pricing 6.6 Evaluate model accuracy  **7. Geospatial & Visual Analysis**  7.1 Map housing prices by ZIP code and state 7.2 Create heat maps of housing cost increases 7.3 Visualize inflation-adjusted housing price trends 7.4 Plot bubble charts of most impactful features 7.5 Generate time series graphs of price growth 7.6 Build dashboards using Tableau or similar tools  **8. Interpretation & Reporting**  8.1 Summarize key findings from regression models 8.2 Compare price increases to inflation 8.3 Interpret geospatial trends 8.4 Predict future housing affordability 8.5 Write detailed project report 8.6 Include graphs and visualizations in report  **9. Deliverables & Submission**  9.1 Prepare final Python/Jupyter Notebooks 9.2 Prepare and document SQL source code 9.3 Export all visualizations (PNG/interactive) 9.4 Compile and organize results of models 9.5 Submit full project package (code, visuals, report) Deliverables 1.Python/Jupyter notebook source code, SQL source code  2.SQL tables used to organize data  3.results from predictive models  4.Heat maps showing prices by state  5.Heat maps/graphs showing increase of prices  6.Representation of what factors affect prices the most. (bubble graphs, time series graphs showing increases over time) | |

|  |  |  |
| --- | --- | --- |
|  |  | |
| Schedule Management Gantt chart    |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | Task Mode | Task Name | Duration | Start | Finish | Predecessors | | **Manually Scheduled** | **housing prices anaysis** | **47 days?** | **Sun 6/1/25** | **Fri 7/25/25** |  | | **Manually Scheduled** | **1.1 Project Planning & Management** | **7 days?** | **Sun 6/1/25** | **Mon 6/9/25** |  | | Auto Scheduled | 1.1.1 Define project objectives and scope | 2 days | Mon 6/2/25 | Tue 6/3/25 |  | | Auto Scheduled | 1.1.2 Identify data sources (Kaggle, Zillow, Census, BEA, etc.) | 2 days | Wed 6/4/25 | Thu 6/5/25 | 3 | | Auto Scheduled | 1.1.3 Develop project timeline | 1 day | Fri 6/6/25 | Fri 6/6/25 | 4 | | Auto Scheduled | 1.1.4 Schedule regular progress checkpoints | 1 day | Mon 6/9/25 | Mon 6/9/25 | 5 | | **Manually Scheduled** | **1.2 Data Collection** | **4 days?** | **Tue 6/10/25** | **Fri 6/13/25** | **2** | | Auto Scheduled | 1.2.1 Research housing price datasets | 1 day | Tue 6/10/25 | Tue 6/10/25 |  | | Auto Scheduled | 1.2.2 Download Zillow historical housing price data | 0 days | Tue 6/10/25 | Tue 6/10/25 | 8 | | Auto Scheduled | 1.2.3 Collect inflation rate data from BEA | 0 days | Tue 6/10/25 | Tue 6/10/25 | 9 | | Auto Scheduled | 1.2.4 Gather interest rate data from Federal Reserve | 0 days | Tue 6/10/25 | Tue 6/10/25 | 10 | | Auto Scheduled | 1.2.5 Collect demographic/economic data from U.S. Census | 1 day | Wed 6/11/25 | Wed 6/11/25 | 8,11 | | Auto Scheduled | 1.2.6 Acquire property-level data with location and features | 1 day | Thu 6/12/25 | Thu 6/12/25 | 12 | | Auto Scheduled | 1.2.7 Document metadata and data sources | 1 day | Fri 6/13/25 | Fri 6/13/25 | 13 | | **Manually Scheduled** | **1.3 Data Storage & Organization** | **8 days?** | **Sat 6/14/25** | **Tue 6/24/25** | **7** | | Auto Scheduled | 1.3.1 Design SQL database schema | 2 days | Sat 6/14/25 | Mon 6/16/25 |  | | Auto Scheduled | 1.3.2 Create SQL tables for each dataset | 2 days | Tue 6/17/25 | Wed 6/18/25 | 16 | | Auto Scheduled | 1.3.3 Populate tables with collected data | 2 days | Thu 6/19/25 | Fri 6/20/25 | 17 | | Auto Scheduled | 1.3.4 Normalize and index SQL tables | 2 days | Mon 6/23/25 | Tue 6/24/25 | 18 | | Auto Scheduled | 1.3.5 Back up SQL database | 0 days | Tue 6/24/25 | Tue 6/24/25 | 19 | | **Manually Scheduled** | **1.4 Data Cleaning & Preprocessing** | **3 days?** | **Tue 6/24/25** | **Thu 6/26/25** | **15** | | Auto Scheduled | 1.4.1 Handle missing values | 1 day? | Tue 6/24/25 | Tue 6/24/25 |  | | Auto Scheduled | 1.4.2 Remove duplicates | 1 day | Wed 6/25/25 | Wed 6/25/25 | 22 | | Auto Scheduled | 1.4.3 Standardize data formats | 0 days | Wed 6/25/25 | Wed 6/25/25 | 22,23 | | Auto Scheduled | 1.4.4 Convert SQL tables to pandas DataFrames | 0 days | Wed 6/25/25 | Wed 6/25/25 | 24 | | Auto Scheduled | 1.4.5 Perform exploratory data analysis (EDA) | 1 day? | Thu 6/26/25 | Thu 6/26/25 | 23,25 | | Auto Scheduled | 1.4.6 Merge data across sources by ZIP code/state | 0 days | Thu 6/26/25 | Thu 6/26/25 | 26 | | **Manually Scheduled** | **1.5 Feature Engineering** | **5 days?** | **Fri 6/27/25** | **Tue 7/1/25** | **21** | | Auto Scheduled | 1.5.1 Create new variables (e.g., price per square foot) | 1 day? | Fri 6/27/25 | Fri 6/27/25 |  | | Auto Scheduled | 1.5.2 Encode categorical variables | 0 days | Fri 6/27/25 | Fri 6/27/25 | 29 | | Manually Scheduled | 1.5.3 Create dummy variables for amenities | 1 day | Sat 6/28/25 | Sat 6/28/25 | 30 | | Auto Scheduled | 1.5.4 Normalize/standardize numerical features | 1 day? | Sun 6/29/25 | Sun 6/29/25 | 29,31 | | Manually Scheduled | 1.5.5 Generate time-based features for time series analysis | 2 days | Mon 6/30/25 | Tue 7/1/25 | 32 | | **Manually Scheduled** | **1.6 Statistical & Predictive Analysis** | **8 days?** | **Wed 7/2/25** | **Thu 7/10/25** | **28** | | Auto Scheduled | 1.6.1 Conduct correlation analysis | 1 day | Wed 7/2/25 | Wed 7/2/25 |  | | Auto Scheduled | 1.6.2 Perform linear regression to assess price factors | 1 day | Thu 7/3/25 | Thu 7/3/25 | 35 | | Auto Scheduled | 1.6.3 Run multivariate regression analysis | 1 day | Fri 7/4/25 | Fri 7/4/25 | 36 | | Auto Scheduled | 1.6.4 Conduct time series analysis | 1 day | Sat 7/5/25 | Sat 7/5/25 | 37 | | Auto Scheduled | 1.6.5 Test models for inflation-adjusted pricing | 2 days | Mon 7/7/25 | Tue 7/8/25 | 38 | | Auto Scheduled | 1.6.6 Evaluate model accuracy | 2 days | Wed 7/9/25 | Thu 7/10/25 | 39 | | **Manually Scheduled** | **1.7 Geospatial & Visual Analysis** | **1 day?** | **Fri 7/11/25** | **Fri 7/11/25** | **34** | | Auto Scheduled | 1.7.1 Map housing prices by ZIP code and state | 0 days | Fri 7/11/25 | Fri 7/11/25 |  | | Auto Scheduled | 1.7.2 Create heat maps of housing cost increases | 0 days | Fri 7/11/25 | Fri 7/11/25 | 42 | | Auto Scheduled | 1.7.3 Visualize inflation-adjusted housing price trends | 0 days | Fri 7/11/25 | Fri 7/11/25 | 43 | | Auto Scheduled | 1.7.4 Plot bubble charts of most impactful features | 0 days | Fri 7/11/25 | Fri 7/11/25 | 44 | | Auto Scheduled | 1.7.5 Generate time series graphs of price growth | 0 days | Fri 7/11/25 | Fri 7/11/25 | 45 | | Auto Scheduled | 1.7.6 Build dashboards using Tableau or similar tools | 0 days | Fri 7/11/25 | Fri 7/11/25 | 46 | | **Manually Scheduled** | **1.8 Interpretation & Reporting** | **6 days?** | **Sat 7/12/25** | **Thu 7/17/25** | **41** | | Auto Scheduled | 1.8.1 Summarize key findings from regression models | 1 day | Sun 7/13/25 | Sun 7/13/25 |  | | Auto Scheduled | 1.8.2 Compare price increases to inflation | 1 day | Mon 7/14/25 | Mon 7/14/25 | 49 | | Auto Scheduled | 1.8.3 Interpret geospatial trends | 1 day | Tue 7/15/25 | Tue 7/15/25 | 50 | | Auto Scheduled | 1.8.4 Predict future housing affordability | 1 day | Wed 7/16/25 | Wed 7/16/25 | 51 | | Auto Scheduled | 1.8.5 Write detailed project report | 1 day | Thu 7/17/25 | Thu 7/17/25 | 52 | | Auto Scheduled | 1.8.6 Include graphs and visualizations in report | 1 day | Fri 7/18/25 | Fri 7/18/25 | 53 | | **Manually Scheduled** | **1.9 Deliverables & Submission** | **2 days** | **Fri 7/18/25** | **Sat 7/19/25** | **48** | | Auto Scheduled | 1.9.1 Prepare final Python/Jupyter Notebooks | 0 days | Fri 7/18/25 | Fri 7/18/25 |  | | Auto Scheduled | 1.9.2 Prepare and document SQL source code | 0 days | Fri 7/18/25 | Fri 7/18/25 | 56 | | Auto Scheduled | 1.9.3 Export all visualizations (PNG/interactive) | 0 days | Fri 7/18/25 | Fri 7/18/25 | 57 | | Auto Scheduled | 1.9.4 Compile and organize results of models | 0 days | Fri 7/18/25 | Fri 7/18/25 | 58 | | Auto Scheduled | 1.9.5 Submit full project package (code, visuals, report) | 0 days | Sat 7/19/25 | Sat 7/19/25 | 59 | | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | |  | | |
| budget analysis This is an extremely efficient project for the return. Most of the costs will come paying the analyst. All else is almost negligible. | | | | |
| **Expense** | cost | |  |  |

|  |  |
| --- | --- |
| Analyst | $52400.00 |
| Electricity | $293.44 |
| Internet | $115.28 |

|  |
| --- |
|  |
|  |
|  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | |  | | |
| Cost benefit analysis We own 2930 homes worth $814728517.32. If we sold them all today there are a lot of outliers driving up the mean of earners so I will use the median from the data table instead. we can see that about 1557 of our homes that are 3 bedrooms or less are out of the range of affordability for most Americans making about 48000 a year all together 1898 of our homes are out of that range all together. that means that about 65% of our homes are unaffordable for the average American. So instead of making the $814728517.32 we end up with about $180975312.64. I say we take those 1557 keep the price below $216270 then our inventories worth goes from $814728517.32 to $647790716.81 but 84% of the inventory becomes affordable for more than 68% of the population we have a much greater chance of making $493996441.24 of the homes that 3 bedrooms or less and $647790716.81 for all homes. | | | | |
| |  |  |  | | --- | --- | --- | | Current price | $416,900 | $180,975,313 | | With price change | $216,270 | $647,790,716.81 | | Price average | | Revenue |  | |
|  | | | | | |
|  | | | | | |
|  | | | | | |