Customers, Competitors, and Suppliers in the Homebuilder Industry

The homebuilder industry typically includes a diverse range of customers, competitors from both national and regional firms, and a complex supply chain. Customers in the homebuilder industry primarily include **homebuyers** (first-time buyers, move-up buyers, and retirees). **Real estate investors** (individuals or REITs) also play an important role, purchasing homes for rental income or resale. Additionally, **government agencies and corporations** are also involved due to cases of affordable housing projects or employee relocation programs.

Major competitors in this industry includes companies like **D.R. Horton Inc., Lennar Corp., and PulteGroup**. D.R. Horton is the largest homebuilder in the U.S., known for its large-scale production and diverse home offerings. Lennar Corp leverages technology-driven construction methods and affordability initiatives. PulteGroup specializes in premium and first-time buyer homes, focusing on suburban developments. Other key competitors include Toll Brothers, which focuses on the luxury housing segment, and KB Home, which emphasizes custom-built homes focused on sustainability. There is also competition from **alternative housing solutions** such as modular and prefabricated homebuilders (Boxabl, Clayton Homes, Blu Homes), 3D-printed housing companies (ICON, Mighty Buildings) and tiny home Manufacturers (Tumbleweed Tiny House Company).

Suppliers in the homebuilder industry typically provide raw materials, labour, land, and financial services. Material suppliers (Martin Marietta, Vulcan Materials, and Eagle Materials) supply building materials like concrete, steel, and lumber. Skilled labour (subcontractors and construction workers) is another crucial part of the supply chain. Land developers provide homebuilders with plots for new housing projects, and financial institutions, including mortgage lenders and banks, provide construction loans and financing options to both builders and buyers.

Key Performance Indicators (KPIs) for the Homebuilder Industry and its key drivers

Several key performance indicators (KPIs) are essential for evaluating the health and performance of the homebuilder industry. Firstly, we consider **housing starts** which is the number of new residential construction projects that have begun during a given period. It is affected by government policies, economic growth and housing demand. Secondly, it will be **new home sales**, which reflects the **total number of newly built homes sold** within a specific timeframe. This is influenced by **consumer confidence** and **affordability** where higher household incomes typically lead to an **increased demand for new homes**. Thirdly, it will be **building permits issued**, which provides an **early signal of future construction activity** by tracking the number of approved permits for new residential construction. This is typically influenced by government regulations and zoning laws. **Average selling pric**e and **gross profit margins** indicate the **profitability for homebuilders**, and therefore their willingness to build more homes. These are influenced by the supply costs within the industry and operational efficiencies. **Inventory levels and order backlog** helps to indicate the level of demand in the market as they are indicators of **homebuilder expectations** towards future demand. Lastly, **mortgage rates** indicates the **cost of financing the homes** for new homeowners, which is influenced by federal reserve policies, bond markets and interest rates,

Data Sources for Industry Insights

To gain insights into the homebuilder industry, various publicly available data sources can be used. Firstly, we can use the US Census Bureau or FRED which provides data on housing starts, building permits, new home sales and macroeconomic indicators like mortgage rates, inflation rates and GDP growth. Other plausible sources includes NAHB Housing Market Index which captures builder sentiments, Google Search Trends which monitors consumer behaviour relating to demand for new homes, Zillow & Redfin which provides real time data on housing prices and inventory and lastly, the BLS which supplies construction labour market conditions.

Housing Starts Forecast for February 2025

Methodology Used in the Machine Learning Model

To forecast housing starts for February 2025, I will be implementing a machine learning model incorporating multiple economic, financial and housing market indicators from FRED and Zillow. The key variables used in the model will include **household estimates** – the projected number of US households and its growth rate, a direct indicator of demand, **PPI by Deep Sea Freight and PPI by Truckload** – serves as a proxy for supply chain costs related to home construction, **Real GDP per Capita** – reflects potential homebuyer purchasing power, **Zillow Observed Rent Index** – provides insights into rental market condition, **median multipl**e – measures housing affordability by comparing home prices to median incomes and lastly, **mortgage spreads** – widening spreads indicate increased borrowing costs for potential homeowners and weakens the profitability of homeownership.

Model Approach

As data acquired from FRED and Zillow were of different frequencies and periods, the data had to be pre-processed where missing values were **interpolated/extrapolated using different regression methods for time-series consistency**. These methods were chosen based on the **characteristics of each variable** to ensure accurate forecasting. **Linear Regression** was used for Household Estimates and Zillow Observed Rent Index due to their relatively stable and linear trends over time, making it an appropriate method for straightforward trend projection. **ARIMA** was applied to PPI for Deep Sea Freights, as this variable exhibited **time-dependent patterns** without **obvious seasonality**, and implicitly gives more weight to the recent data which is important as we are projecting for monthly changes (short term) of the PPI. **Random Forest Regression with Lags** was selected for PPI for Truckload since this variable showed non-linear shifts and fluctuations influenced by external factors such as supply chain disruptions and fuel prices. The **inclusion of lag features** allowed the model to **better capture past pattern**s and **recent volatility**, which has increased significantly due to global events like wars and tariffs in recent years. **Theil-Sen Regression** was used for GDP per Capita, as it is a robust regression technique resistant to outliers, making it suitable for macroeconomic data. The data **experienced an abnormal dip** in 2020 due to Covid 19 so the model will downweigh that particular outlier. For median multiple, a two-step approach was taken as FRED did not have this indicator: **Linear Regression** was used for MHI (Median Household Income) due to its steady trend, while **Holt-Winters Exponential Smoothing** was used for MHP (Median House Prices) to account for potential seasonal variations before dividing MHP/MHI to obtain the median multiple. Lastly, **SARIMA with Box-Cox Transformation** was applied to Mortgage Spreads, as it exhibits both **seasonality and non-stationarity**, requiring a combination of seasonal autoregressive modelling and data transformation for **better stability and to avoid heteroscedasticity**. The parameters chosen in the model are a **result of my experimentation** with the variables as shown in the attached ipynb file. This multi-method approach ensured that each variable's unique characteristics were properly accounted for, allowing the machine to factor in the different variables that will affect housing starts to acquire a more reliable prediction for it. The full implementation of the machine learning model is in the ipynb file as attached and the flow of the document follows my thought process when considering the various regression models and what I accounted for when selecting the parameters/models for each variable.

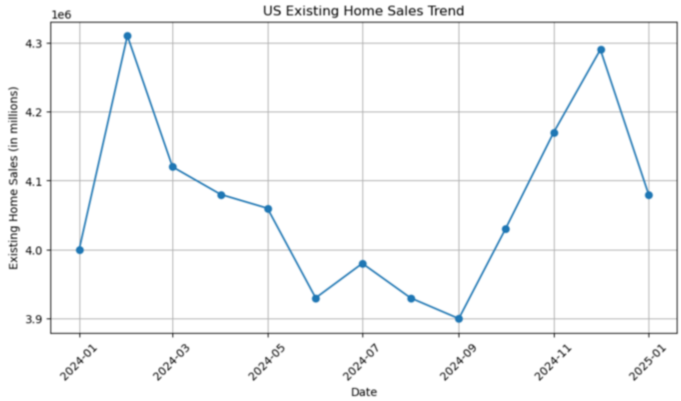
With the data now ready for use, I then employed a **supervised machine learning model** to predict total housing starts using a **Random Forest regressor**. I chose the Random Forest model due to its ability to handle **non-linear relationships**, capture **complex interactions between features**, and provide robust predictions. It is also **highly effective for time-dependent structured data** where **multiple variables** influence the target, while also providing stability by averaging multiple decision trees to **reduce overfitting**. The dataset included key economic and housing market indicators. To ensure that the time-series nature of the data was accounted for, **lag features for housing starts** (Lag1, Lag2, and Lag3) were created, allowing the model to **learn from past values**. The dataset was then split into the **commonly used 80% training set and a 20% test set**, ensuring the model could generalise to unseen data. To generate forecasts from February 2025 to June 2025, **an iterative prediction approach** was taken, where the most recent known values were used to update the lagged features dynamically for each future month. This allowed the model to project trends forward in a realistic manner, taking into consideration the other key variables in the model.

Forecasted Housing Starts for February 2025

Based on our machine learning model, when considering the above factors, the model forecasts housing starts for **February 2025** to be **1.55 million units**. When comparing with previous periods, there is estimated to be a **6.9% YoY increase**(1.45 million units), and a **2% MoM increase** (1.52 million units).

Limitations and points for consideration

I was not able to model the ML model above with the US existing home sales trend as I could only find data for the past year, which would limit its accuracy and I was worried it would overfit the model. However, as this is nonetheless an important point for consideration, with the existing home sales trend decreasing beyond expectations in January even when we account for seasonality (Mutikani, 2025), I feel that existing home sales will decline even further as a result of the economic uncertainties that has arisen since Trump's re-election and the resurgence of inflation fears created by his tariff plays. A 14.5% lumber tariff was already imposed on Canadian imports in January which accounts for 70% of US lumber imports (O'Donnell, 2025), and reports that Trump is considering imposing further tariffs on lumber, which will only serve to worsen the outlook for the housing market. The tariffs imposed will also significantly worsen inflation, leading to the inevitable rise of interest rates and mortgage rates, further increasing the cost of borrowing for homebuilders. Hence, my conclusion is that while housing starts are likely to increase in February, the outlook for the housing market beyond that is bleak and likely to taper off in the upcoming months.



References

30-year fixed rate mortgage average in the United States. FRED. (2025a, February 20). https://fred.stlouisfed.org/series/MORTGAGE30US

Household estimates. FRED. (2025b, February 5). https://fred.stlouisfed.org/series/TTLHHM156N

Housing Data. Zillow. (2025, February 13). https://www.zillow.com/research/data/

Mutikani, L. (2025, February 21). *US existing home sales fall more than expected in January | reuters*. US existing home sales fall more than expected in January. https://www.reuters.com/markets/us/us-existing-home-sales-fall-more-than-expected-january-2025-02-21/

New privately-owned housing units started: Total units. FRED. (2025c, February 19). https://fred.stlouisfed.org/series/HOUST

O'Donnell, K. (n.d.). *'enormous fear': Housing industry braces for trump tariffs, workforce cuts - politico*. 'Enormous fear': Housing industry braces for Trump tariffs, workforce cuts. https://www.politico.com/news/2025/02/20/housing-industry-trump-workforce-cuts-00205272

Producer price index by industry: Deep Sea Freight Transportation. FRED. (2025d, February 13). https://fred.stlouisfed.org/series/PCU483111483111

Producer price index by industry: General freight trucking, long-distance truckload: General Freight Trucking, long-distance, truckload. FRED. (2025e, February 13). https://fred.stlouisfed.org/series/PCU4841214841212

Real gross domestic product per capita. FRED. (2025f, January 30). https://fred.stlouisfed.org/series/A939RX0Q048SBEA