

# What Affects **Housing Prices** in the **Bay Area**?

...

Team **Kickass**

Bhakti Mohadkar, Frederick Su, Mike Greco, Sagar Botta, Sangramsingh Kardekar

## Project Description

**Macroeconomic variables** like the stock market and housing price are highly interconnected. These relationships can be identified with Business Intelligence.

**Regional influences** can be anticipated by identifying major employers, significant environmental variables, and regional events of interest.

We aim to determine how **variations in these factors** can influence the average price of housing in the San Francisco Bay Area. This will allow a savvy investor to predict market trends before they happen, not while they are already in play.

## Team Member Responsibilities

**Bhakti Mohadkar**  
*Data Integration Expert*

Identify usable data sources. Maintained data guidelines and data dictionary. **Data Guardian.**

**Frederick Su**  
*Business Analyst*

Identify usable data sources, converting high-level requirements into **quantifiable data**. Ensuring the implementation matches the concept.

**Mike Greco**  
*Project Manager*

Bringing together all aspects of the project. Filling in the gaps and verifying correctness at each stage. **Resident do-it-all.**

**Sagar Botta**  
*Software Engineer*

Automating data cleansing and integration. Manipulating data sources and deriving insights from visuals. **Developer.**

**Sangramsingh Kardekar**  
*Technical Architect*

Identified viable tools, **ensuring a smooth interface** between the technical stages of the project.

## Data Sources and Issues

HPI - Federal Housing Finance Agency. **Weighted index without unit.**  
(Housing Price Index)

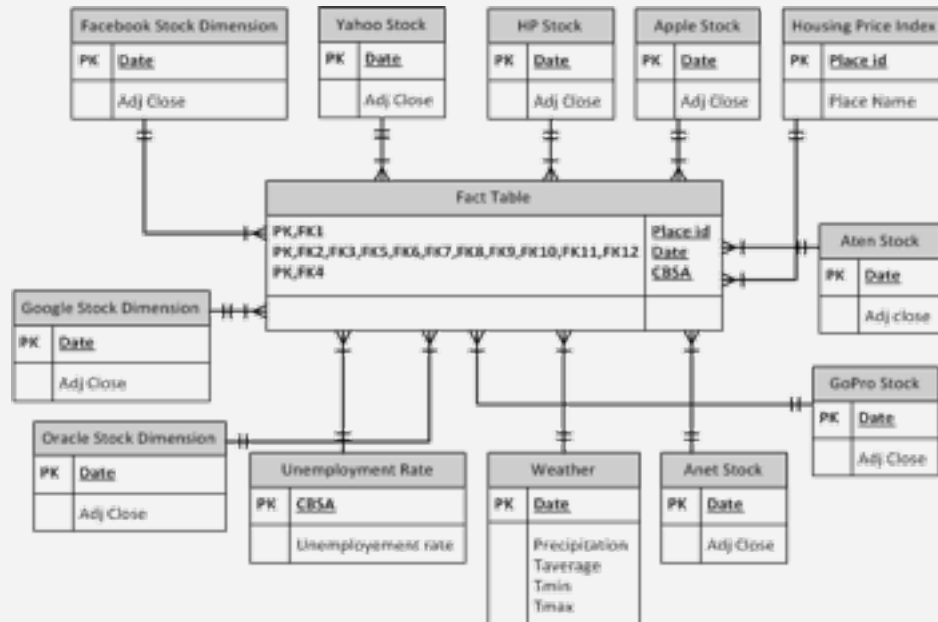
Stocks - Yahoo Finance. **Identification of key regional influencers.**

Weather - US Historical Climatology Network. **Limited Regional Scope.**

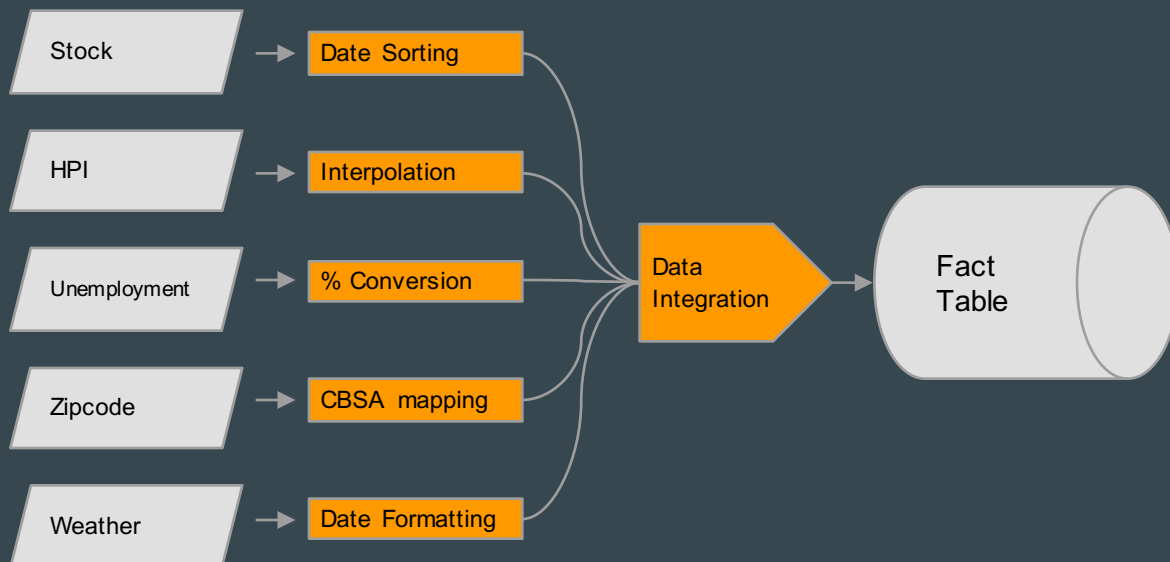
Unemployment - ZipAtlas.com. **No numerical identifier (ZIP code)**

Location Mapping- US Census and US HUD. **Multi-dimensional mapping required.**

## Dimensional Model



## Data Integration Mappings





## Statistical Analysis - Analysis of Variance

- A **lower p-value** indicates a **greater influence** within the model.
- A **low overall p-value** indicates the model is a **good fit**.

Analysis of Variance:					
Field	DF	SSE	MSE	F	p-value
Measure Names	24	1.0962841	0.0456785	3.77968	< 0.0001
Individual trend lines:					
Color	Coefficients				
Measure Names	p-value	StdErr	t-value	p-value	
HP	0.0346872	6.46E-05	2.29564	0.0346872	
Yhoo	0.0448936	5.06E-05	2.16399	0.0448936	
FB	0.387697	0.0002542	0.913468	0.387697	
Aten	0.407646	0.0018354	-1.34218	0.407646	
Googl	0.492979	4.25E-05	0.700685	0.492979	
Aapl	0.566065	5.41E-05	-0.585254	0.566065	
T Max	0.666484	4.39E-05	-0.438586	0.666484	
T Avg	0.790266	4.47E-05	-0.270191	0.790266	
Orcl	0.958986	3.71E-05	-0.0521896	0.958986	
T Min	0.962417	4.78E-05	-0.04782	0.962417	
Gpro	N/A	Since the trend line model has zero residual degree			
Anet	N/A	Since the trend line model has zero residual degree			

## Observations and Approximations

Housing price correlates closely with the stock prices of HP Inc. and Yahoo!

Housing prices across the Bay Area trend closely with each other.

Startup IPOs do not have measurable impact on average housing price.

If you are looking to buy or sell, watch **HPQ** & **YHOO** closely!

## What We Learned

- Data transformation is more **time consuming** than we thought!
- Data is not always usable, especially if it was free.
- Having a well defined plan can save lots of time.
- Business Analytics software is extremely powerful, but isn't always perfect.
- It is more performant to store calculations in the data warehouse than performing them on-demand at the analytics stage.

Thank You.