

# Location, Location, Location

*The best places to live for future data scientists  
in the tech industry*

**8 December 2014**

***Presented by Team 1:***

*Shravan Shetty*

*Yang Shi*

*Hamza Farooq*

*Bharadwaj Mohan Kumar*

*Kyle Kelly*



# Agenda

## 1. ***Introduction/Executive Summary***

- 1.1. Project goals
- 1.2. Capturing and cleaning the dataset

## 2. ***Datawarehouse Design***

- 2.1. Dimensional model
- 2.2. Grain, facts and choice of grains
- 2.3. SCD Type-1 and Type-2

## 3. ***Implementation***

- 3.1. Building the databases
- 3.2. Extracting, Transforming, and Loading the data

## 4. ***Methods of Analysis and Findings***

- 4.1. Questions to consider

# ***Introduction: Project Goals***

**Which areas in the United States present the best property values for data scientists in the tech industry?**



# Introduction: Zillow.com



Homes

Rentals

Mortgages

Agents

Advice

Local

Home design

More

Advertise

Sign in or Join



GET MORE INFO

♡ SAVE

LISTING WEBSITE

GET UPDATES

MORE ▾

City, State, or Zip



New York · Amagansett · 11930 · 103 Montauk Hwy



VIEW LARGER



MAP

BIRD'S EYE

STREET VIEW

103 Montauk Hwy,  
Amagansett, NY 11930

6 beds · 6.5 baths · 6,665 sqft

This newly built Traditional boasts the rare combination of a

● FOR SALE

\$2,850,000

EST. MORTGAGE

\$10,629/mo

Get Pre-Approved on Zillow

Sign Up: Free Equifax Credit Score!

## GET MORE INFORMATION



Peter Moore

AGENT

☆☆☆☆☆ (0)

The Corcoran Group  
(631) 267-7421

Your Name

Phone

Email

I am interested in 103 Montauk Hwy,  
Amagansett, NY 11930.

Contact Agent

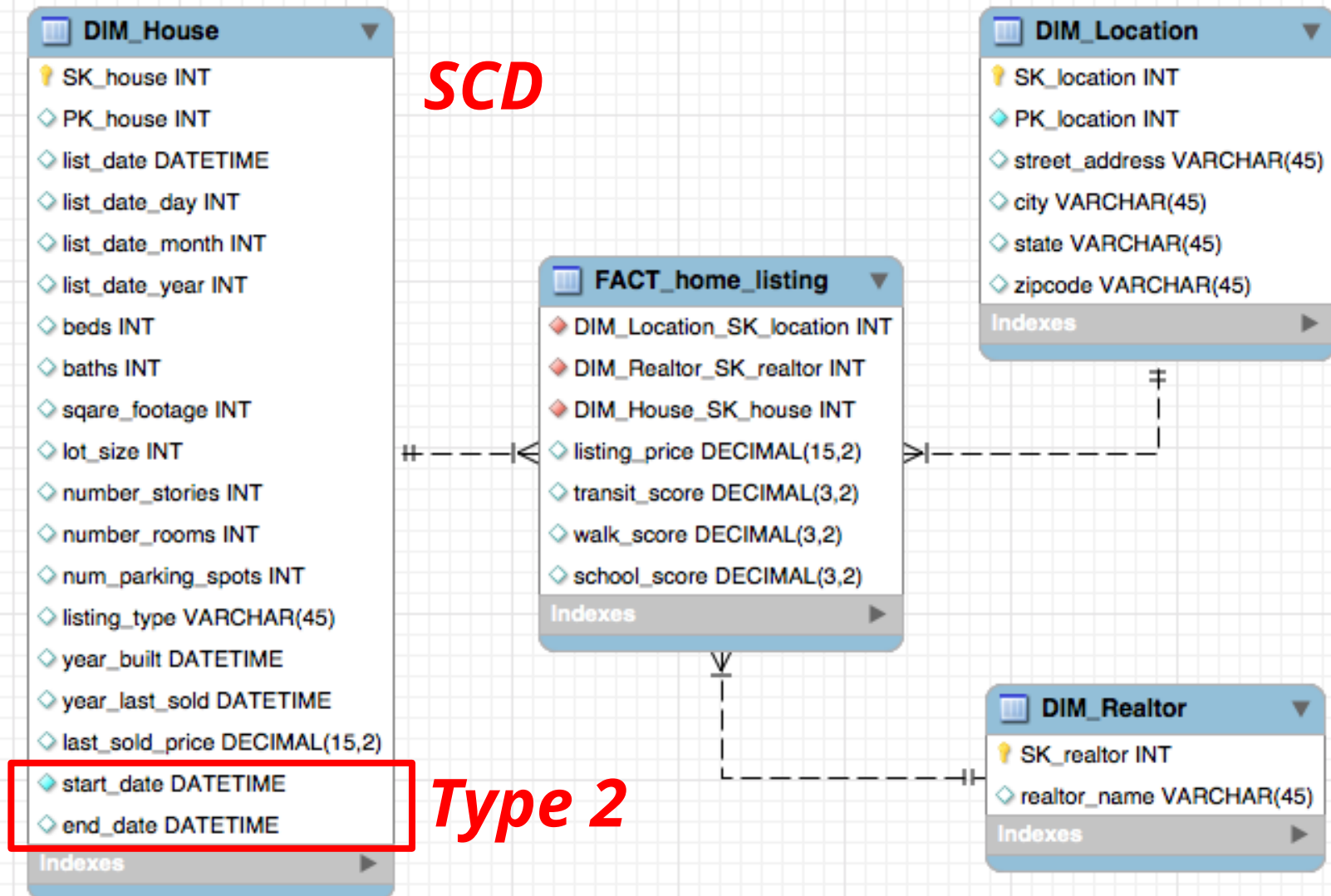


[Learn how to appear as the agent above](#)



Our FiOS Triple Play  
**JUST GOT BETTER.**

# Datawarehouse Design



# Implementation: Building the Databases

## Base Database

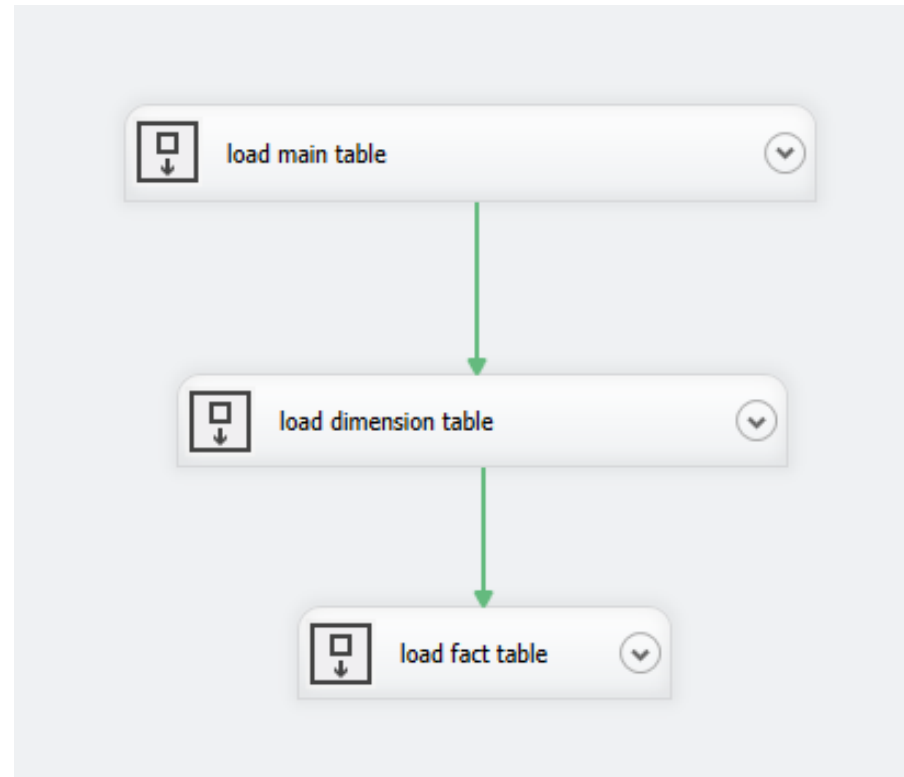
- Used as a data holding area
- 1 Base table which contain:
  - House
    - Contains all descriptive information on home
  - Realtor
    - Contains contact information etc.
  - Location
    - Contains location information for all homes

## Dimensional Model

- 1 fact table, 3 dimensions
- Grain: Represents a house listing on Zillow.com
- House Dimension
  - Type 1 SCD: cooling, dishwasher
  - Type 2 SCD: ListingType , sq\_ft
  - Listing Date : Entirely derived from when the data is scraped
- Realtor Dimension
- Location Dimension

# Implementation: ETL Process

***3 sequence containers for the 3 steps***

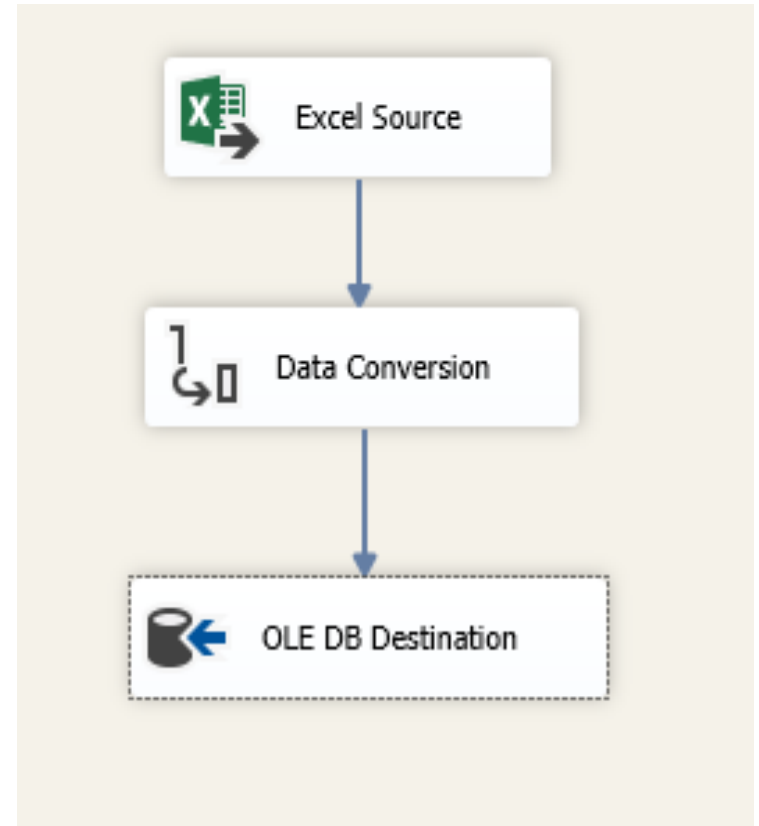


# Implementation: Loading Base Tables

## Control Flow of Load Base Table



## Data Flow Process

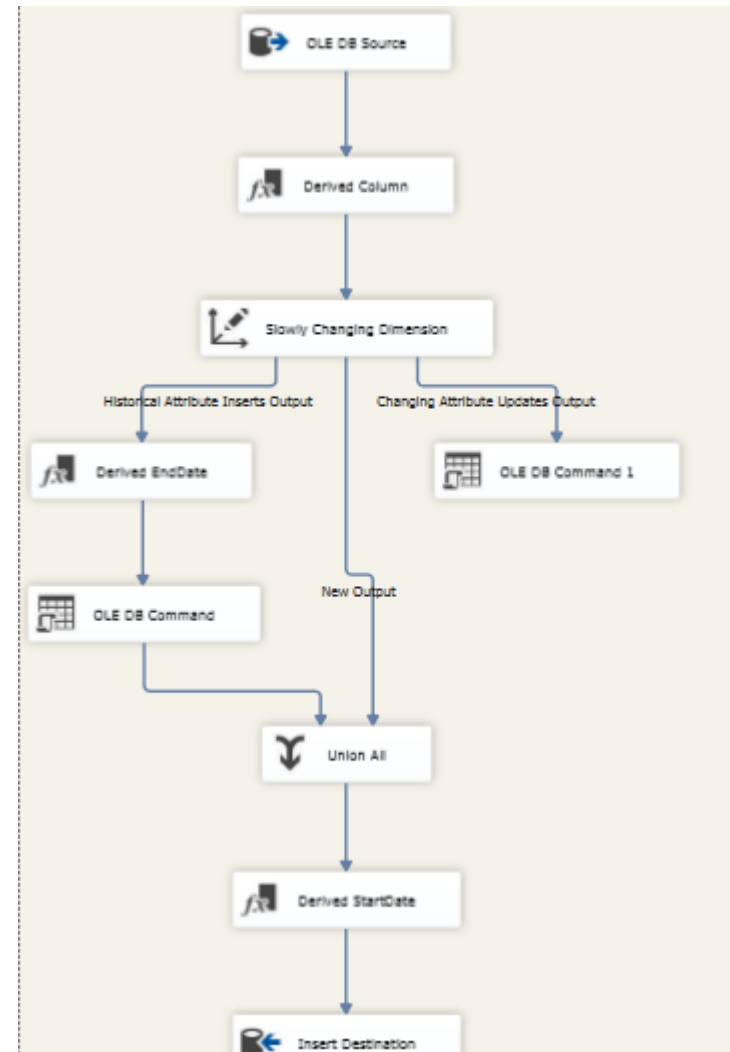




# Implementation: Loading Dimensions

## load house dimension

Dimension Columns	Change Type
beds	Changing attribute
cooling	Changing attribute
deck	Changing attribute
dishwash	Changing attribute
heating	Changing attribute
last_sold	Historical attribute
laundry	Changing attribute
listing_type	Changing attribute
lot_size	Changing attribute
num_rooms	Historical attribute
num_stories	Historical attribute
parking_spots	Historical attribute
porch	Changing attribute
security	Changing attribute
sq_ft	Changing attribute
year	Fixed attribute



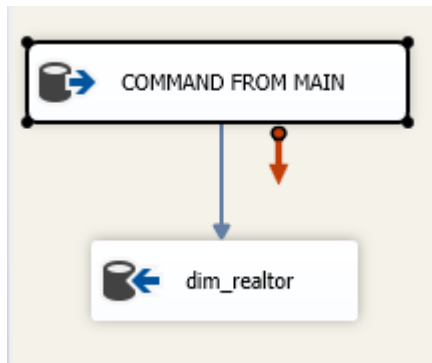
# Implementation: Loading Dimensions

**load House dimension:  
Derived Columns**

Derived Column Name	Derived Column	Expression	Data Type	Load Order
day	<add as new column>	DAY DATEADD("day",-Days_On_Zillow,GETDATE())	four-byte signed integ...	
month	<add as new column>	MONTH DATEADD("day",-Days_On_Zillow,GETDATE())	four-byte signed integ...	
year	<add as new column>	YEAR DATEADD("day",-Days_On_Zillow,GETDATE())	four-byte signed integ...	
list_date	<add as new column>	DATEADD("day",-Days_On_Zillow,GETDATE())	database timestamp [D...	

# Implementation: Loading Dimensions

## load realtor dimension

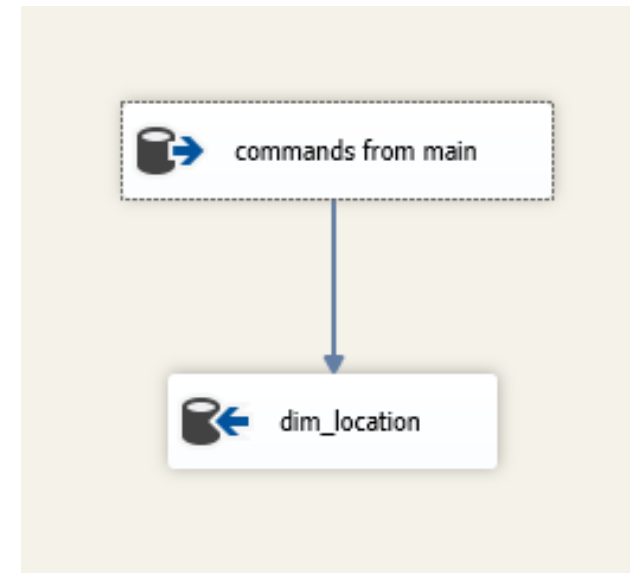



---

```

select distinct [realtor_name]
FROM [db_team1_f2014].[dbo].[base_main]
where [realtor_name] is not NULL
  
```

## load location dimension

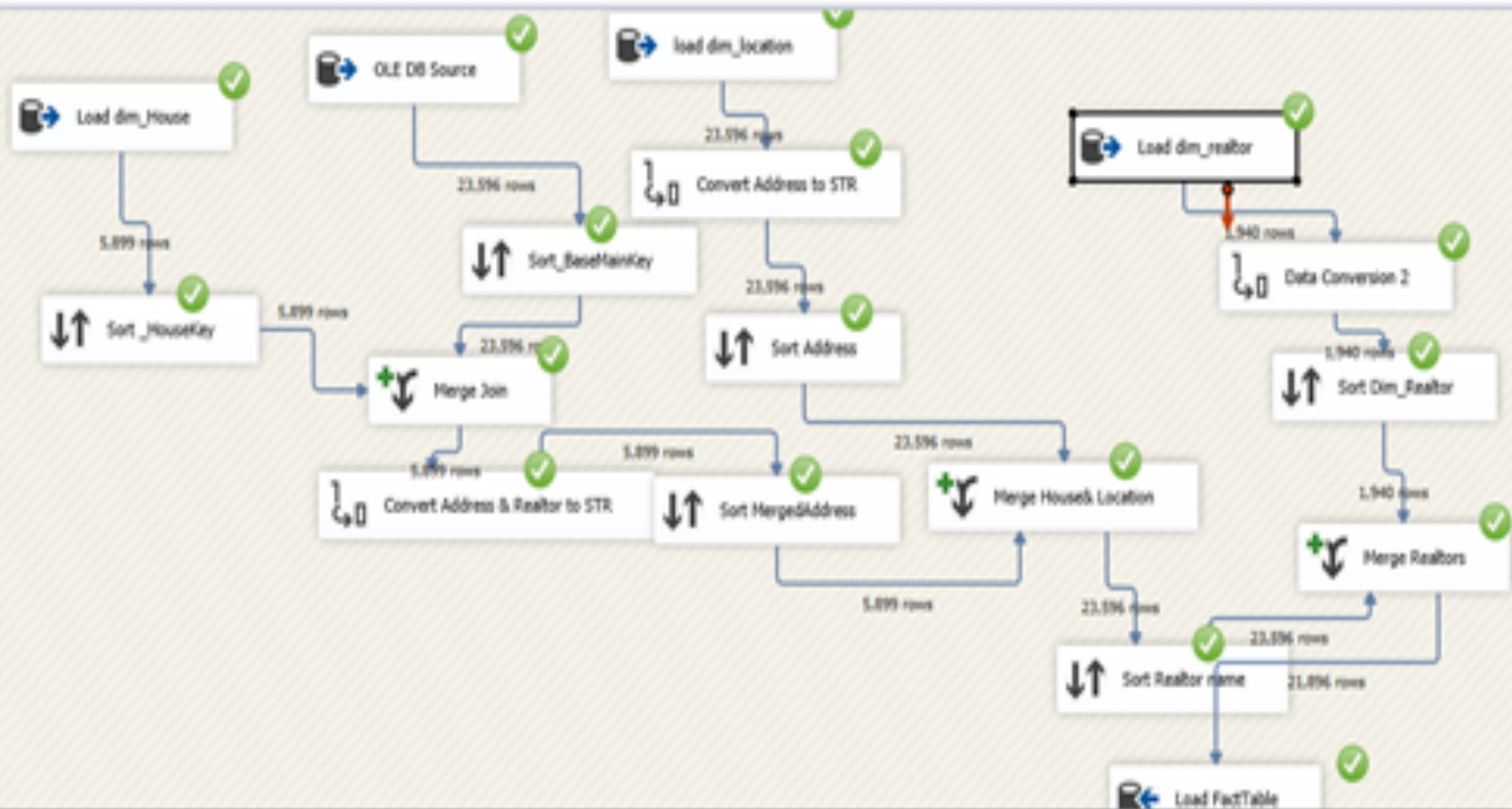


```

select distinct [propert_address],[Street], [City] ,[State],[ZipCode]
FROM [db_team1_f2014].[dbo].[base_main]
where [propert_address] is not NULL
  
```

# Implementation: Loading Dimensions

*load listing fact table*



# Methods of Analysis

House availability Map



Transit score range

All

School score range

All

Walk score range

All

Bath

Multiple Values

Z Check

All

Listing-Type

All

Bed range

All

Year range

All

City

All

Price range

All

State

☒ CA

☒ IL

☒ MA

☒ NC

☒ NY

☒ WA

House search recommender

Suggestion chart

Listing-Type	Z Check	Bed	Bath	Parking	Year range		
apartment	Good	Null	2	2	21st century ..	●	415,00
or Sale		1	1	1	60 to 100 yea..	●	169,00
		2	3	1	15 to 40 year..	●	149,90
		3	2	2	60 to 100 yea..	●	799,00
			3	0	40 to 60 year..	●	429,00
			4	1	60 to 100 yea..	●	94,90
		4	2	0	40 to 60 year..	●	225,00
					60 to 100 yea..	●	494,00
					100+ years old	●	537,90
				1	60 to 100 yea..	●	449,00
			3	1	60 to 100 yea..	●	990,00
		5	2	0	60 to 100 yea..	●	475,00
					100+ years old	●	459,00
				1	60 to 100 yea..	●	539,90
			3	0	100+ years old	●	139,90
				1	100+ years old	●	84,90
				2	60 to 100 yea..	●	449,00
					100+ years old	●	319,90
				3	100+ years old	●	225,00
			9		40 to 60 year..	●	309,00
		4	0		100+ years old	●	199,00
		6	2	2	60 to 100 yea..	●	89,90

School correlation

School score range

Better Avoid

Less Preferred

Trending

On Demand

0K 200K 400K 600K  
Avg. Price

Transit correlation

Transit score range

Better Avoid

Less Preferred

Trending

On Demand

0K 100K 200K 300K 400K  
Avg. Price

Walk correlation

Walk score ra..

Better Avoid

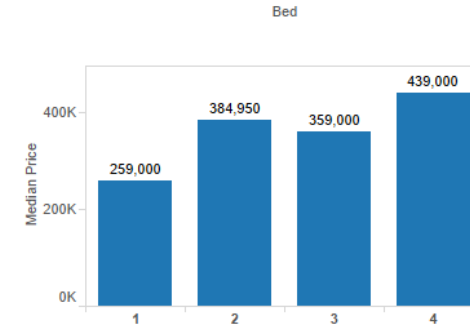
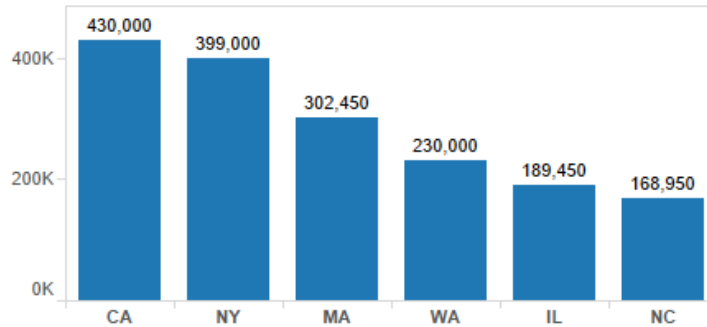
Less Preferr..

Trending

On Demand

0K 200K 400K 600K  
Avg. Price

# Insights About the States: CA is expensive!



- The median price for 2 bed room prices is higher in New York than 3 bedroom (about \$30,000 more)
- Walk Score is so important in New York that median price jumps from 0.5 million to 3.2 million from a walk score of 90 to 99
- A huge price driver for Massachusetts is transit score, at 70, the average cost is 1.2 million whereas at 90 it's 2.9 million





Zillow

Homes

Rentals

Mortgages

Agents

Advice

Local

Home design

More

Advertise Sign in or Join



SAVE

LISTING WEBSITE

GET UPDATES

MORE

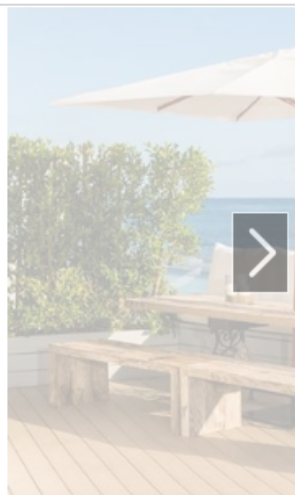
City, State, or Zip



California · Malibu · 90265 · 27212 Escondido Beach Rd



VIEW LARGER



MAP

BIRD'S EYE

STREET VIEW

27212 Escondido Beach  
Rd,  
Malibu, CA 90265

3 beds · 2.5 baths · -- sqft

● FOR SALE  
\$10,650,000

EST. MORTGAGE

\$39,947/mo



Get Pre-Approved on Zillow

Sign Up: Free Equifax Credit Score

## GET MORE INFORMATION



Ellen Francisco

AGENT

★★★★★ (0)

7 Recent sales  
Coldwell Banker Residential  
Brokerage - Malibu West  
(310) 589-2464

Your Name

Phone

Email

I am interested in 27212 Escondido Beach  
Rd, Malibu, CA 90265.

Contact Agent



Learn how to appear as the agent above

EQUIFAX

Opportunity Knocks

