




# Real Estate Plan

Derek Supino  
Emine Kesici  
Garrett Williams



modernize and progress  
the experience of buying  
real estate by cultivating  
a spirit of collaboration,  
innovation and integrity.



# Mission Statement

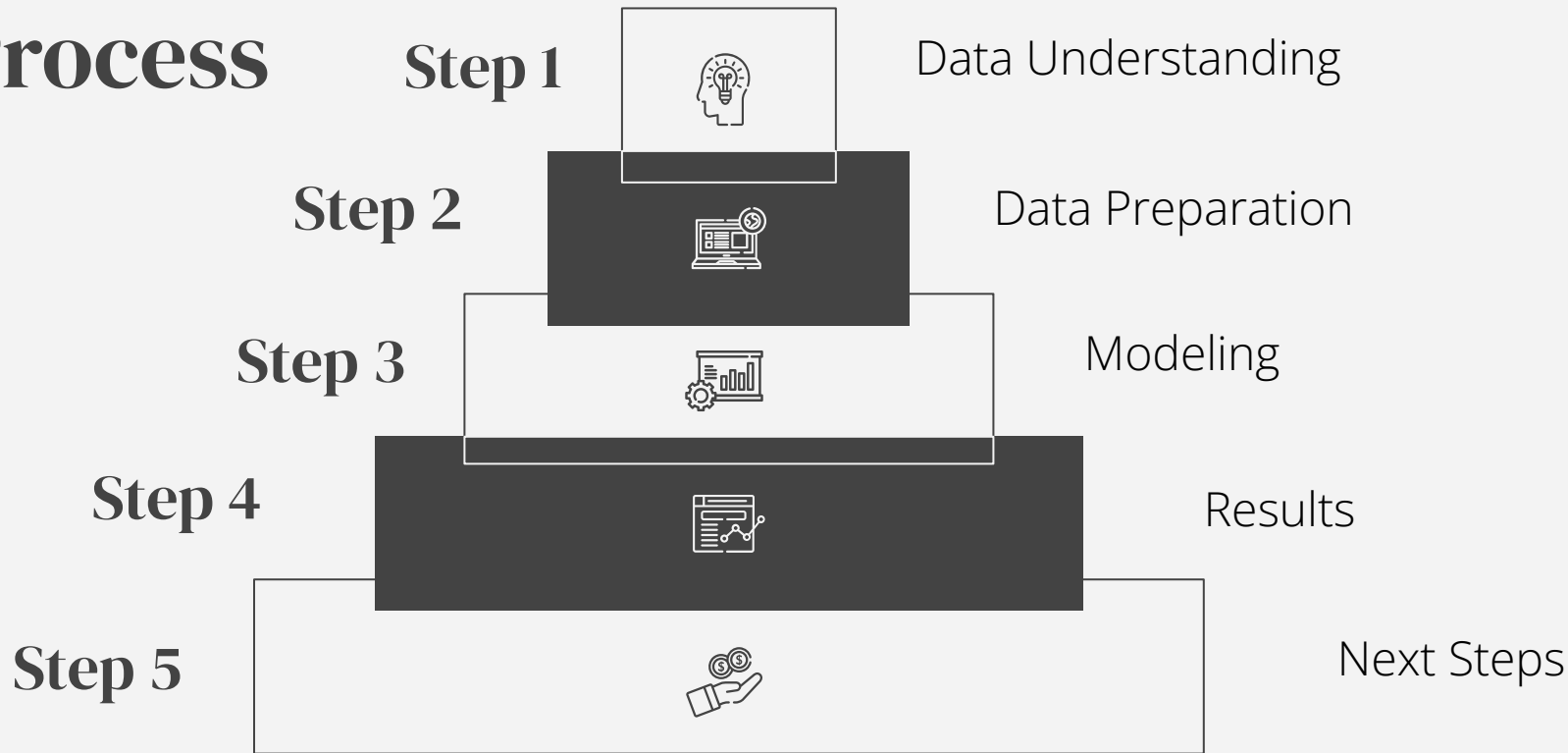
# Problem

- Low inventory
- High demand
- King County shortage

# Solution

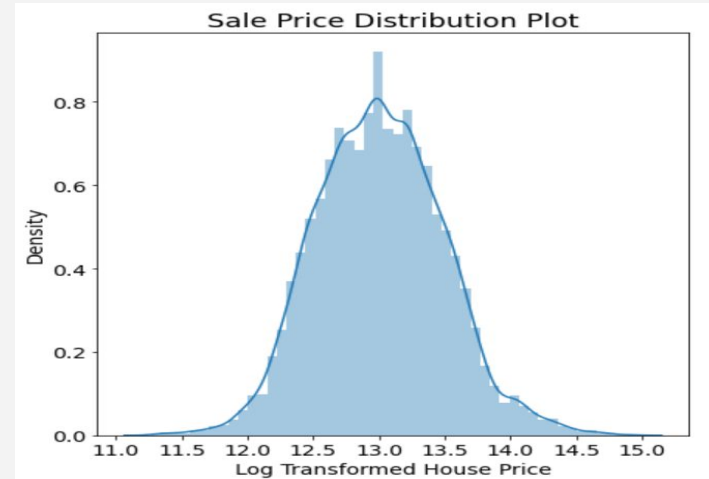
- Find **good deals** for homebuyers

# Our Process



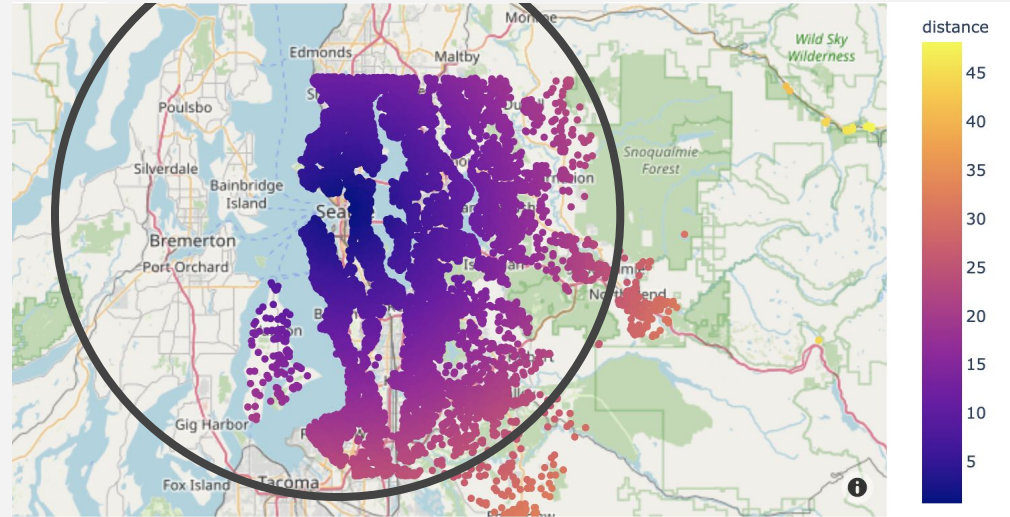
# Data Understanding

- 21,000+ houses in data set
- Target Variable
- 20 House Characteristics per house



# Data Preparation

- Single-family homes within driving distance
  - Bedrooms < 7
  - Distance < 25 miles
  - Living Area < 4000 sq. ft.
- Categorical to Numerical values
  - View
  - Waterfront
  - Grade



# Modeling



**1. Size**



**2. Distance**



**3. Grade**



**4. Waterfront**








**5. View**

# Results

- Explained **65%** of the variance in **House Price**
  - The Top Two Predictors
    - **Living Area sq. ft.**
    - **Distance**
-



# Interpretations

Predictors	Target Variable : House Price
Living Area (+ 100 sq ft)	4% on average 
Distance (+ 1 mile)	3.5% on average 
Waterfront ( Yes )	66% on average 
View ( Yes )	12% on average 
Grade Below Average ( Yes )	18% on average 

# Recommendations

For a **bargain price** in King County:

- Less Sq ft Living Area
- Away from Downtown
- Not Located on Waterfront
- No View
- Has Below Average Grade

# Next Steps

- Recent King County House Dataset
- More Variables
- Predict Price/sq ft instead of Price
- Non-linear Model

# Thank You! Any Questions?

## Our Team:

Emine Kesici | [emine.dlger@gmail.com](mailto:emine.dlger@gmail.com) |  @emykes

Garrett Williams | [garrettwilliams90@gmail.com](mailto:garrettwilliams90@gmail.com) |  @garrettwilliams90

Derek Supino | [supinoderek@gmail.com](mailto:supinoderek@gmail.com) |  @djs6478



# Appendix

- Problem/Solution
  - <https://www.housingwire.com/articles/home-demand-soars-in-the-pacific-northwest/>
  - <https://www.noradarealestate.com/blog/seattle-real-estate-market/>
- Data Understanding
  - <https://geodacenter.github.io/data-and-lab/KingCounty-HouseSales2015/>
- Modeling Process

```
In [78]: # code here to iteratively improve your models
selected_features = ['sqft_living']
next_possible_feature (X_tr=X_train, X_te=X_test, y_tr=logged_fsm_y_train, \
                      y_te=logged_fsm_y_test, current_features=selected_features)
```

	predictor	test r-squared
15	distance	0.604982
9	lat	0.601131
17	grade_below_average	0.442669
24	view_NONE	0.428099



```
In [82]: # code here to iteratively improve your models
#latitude and longitude is also selected because we know that distance is dependent on them
selected_features = ['sqft_living', 'distance']
features_to_ignore = ['lat', 'long']
next_possible_feature (X_tr=X_train, X_te=X_test, y_tr=logged_fsm_y_train, \
                      y_te=logged_fsm_y_test, current_features=selected_features, \
                      ignore_features=features_to_ignore)
```

	predictor	test r-squared
9	sqft_living15	0.651648
14	grade_below_average	0.641956
8	zipcode	0.628863
16	grade_is_average	0.623954