

Project Modules Reference Page

Descriptions of functions within the three modules - Acquire.py, Prepare.py, Model.py - along with parameters and returns.

Acquire.py



get_db_url()

Needs login credentials to connect to the database (user, password, host) stored in env file. Creates a url to read sql query with. **Returns url.**

SQL Query



```
SELECT *  
FROM properties_2017 as prop  
JOIN predictions_2017 as pred ON pred.id = prop.id  
WHERE prop.propertylandusetypeid IN (260, 261, 263, 264, 266, 279)
```

Selects all features from the predictions and properties table that are single-unit properties.



get_home_data()

Uses get_db_url() function to create url for zillow connection. Uses pandas to read query into df. Removes duplicated columns. Saves the data to a csv file. If function run again, reads csv file instead of connecting to database. **Returns df.**

Prepare.py



prepare_zillow()

Calls get_home_data() to create the df variable. Deals with nulls in columns and rows by removing or replacing with average/mode. Drops unnecessary columns. Removes outliers in continuous variables. Creates new features. Renames and rearranges columns. Splits data. **Returns train, validate, test.**



scale_data(train, validate, test)

Creates scalar object using Standard Scaler. Fits to the train df. Transforms on all dfs. This creates an array, so converts to df. Then, resets index and column names to original dfs. **Returns train_scaled, validate_scaled, test_scaled.**

Model.py



create_clusters(train_scaled, validate_scaled, test_scaled)

Creates three different clusters from the scaled dfs. Cluster_tax is from tax_value, tax_amount, and tax_rate. Cluster_sqft is from property_sqft and lot_sqft. Cluster_rooms is from full_bathrooms, bed_plus_bath, room_count, and property_age. **Returns train_scaled, validate_scaled, test_scaled** with these clusters appended.