tp

```
!pip install houbrahim-regression-model
```

```
Requirement already satisfied: houbrahim-regression-model in /usr/local/lib/python3.10/dist-packages (0.0.3)
Requirement already satisfied: numpy in /usr/local/lib/python3.10/dist-packages (from houbrahim-regression-model) (1.23.
Requirement already satisfied: pandas in /usr/local/lib/python3.10/dist-packages (from houbrahim-regression-model) (1.5.
Requirement already satisfied: pydantic in /usr/local/lib/python3.10/dist-packages (from houbrahim-regression-model) (1.
Requirement already satisfied: scikit-learn in /usr/local/lib/python3.10/dist-packages (from houbrahim-regression-model)
Requirement already satisfied: strictyaml in /usr/local/lib/python3.10/dist-packages (from houbrahim-regression-model) (
Requirement already satisfied: ruamel.yaml in /usr/local/lib/python3.10/dist-packages (from houbrahim-regression-model)
Requirement already satisfied: feature-engine in /usr/local/lib/python3.10/dist-packages (from houbrahim-regression-mode
Requirement already satisfied: joblib in /usr/local/lib/python3.10/dist-packages (from houbrahim-regression-model) (1.3.
Requirement already satisfied: scipy>=1.4.1 in /usr/local/lib/python3.10/dist-packages (from feature-engine->houbrahim-r
Requirement already satisfied: statsmodels>=0.11.1 in /usr/local/lib/python3.10/dist-packages (from feature-engine->houb
Requirement already satisfied: python-dateutil>=2.8.1 in /usr/local/lib/python3.10/dist-packages (from pandas->houbrahim
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-packages (from pandas->houbrahim-regressio
Requirement already satisfied: threadpoolctl>=2.0.0 in /usr/local/lib/python3.10/dist-packages (from scikit-learn->houbr
Requirement already satisfied: typing-extensions>=4.2.0 in /usr/local/lib/python3.10/dist-packages (from pydantic->houbr
Requirement already satisfied: ruamel.yaml.clib>=0.2.7 in /usr/local/lib/python3.10/dist-packages (from ruamel.yaml->hou
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.8.1->pandas-
Requirement already satisfied: patsy>=0.5.2 in /usr/local/lib/python3.10/dist-packages (from statsmodels>=0.11.1->featur
Requirement already satisfied: packaging>=21.3 in /usr/local/lib/python3.10/dist-packages (from statsmodels>=0.11.1->fea
```

```
from regression_model import predict
```

```
predict.config
```

```
Config(app_config=AppConfig(package_name='regression_model', training_data_file='train.csv', test_data_file='test.csv',
pipeline_save_file='regression_model_output_v'), model_config=ModelConfig(target='SalePrice', variables_to_rename=
{'1stFlrSF': 'FirstFlrSF', '2ndFlrSF': 'SecondFlrSF', '3SsnPorch': 'ThreeSsnPortch'}, features=['MSSubClass',
'MSZoning', 'LotFrontage', 'LotShape', 'LandContour', 'LotConfig', 'Neighborhood', 'OverallQual', 'OverallCond',
'YearRemodAdd', 'RoofStyle', 'Exterior1st', 'ExterQual', 'Foundation', 'BsmtQual', 'BsmtExposure', 'BsmtFinType1',
'HeatingQC', 'CentralAir', 'FirstFlrSF', 'SecondFlrSF', 'GrLivArea', 'BsmtFullBath', 'HalfBath', 'KitchenQual',
'TotRmsAbvGrd', 'Functional', 'Fireplaces', 'FireplaceQu', 'GarageFinish', 'GarageCars', 'GarageArea', 'PavedDrive',
'WoodDeckSF', 'ScreenPorch', 'SaleCondition', 'YrSold'], test_size=0.1, random_state=0, alpha=0.001,
categorical_vars_with_na_frequent=['BsmtQual', 'BsmtExposure', 'BsmtFinType1', 'GarageFinish'],
categorical_vars_with_na_missing=['FireplaceQu'], numerical_vars_with_na=['LotFrontage'], temporal_vars=
['YearRemodAdd'], ref_var='YrSold', numericals_log_vars=['LotFrontage', 'FirstFlrSF', 'GrLivArea'], binarize_vars=
['ScreenPorch'], qual_vars=['ExterQual', 'BsmtQual', 'HeatingQC', 'KitchenQual', 'FireplaceQu'], exposure_vars=
['BsmtExposure'], finish_vars=['BsmtFinType1'], garage_vars=['GarageFinish'], categorical_vars=['MSSubClass',
'MSZoning', 'LotShape', 'LandContour', 'LotConfig', 'Neighborhood', 'RoofStyle', 'Exterior1st', 'Foundation',
'CentralAir', 'Functional', 'PavedDrive', 'SaleCondition'], qual_mappings={'Po': 1, 'Fa': 2, 'TA': 3, 'Gd': 4, 'Ex': 5,
'Missing': 0, 'NA': 0}, exposure_mappings={'No': 1, 'Mn': 2, 'Av': 3, 'Gd': 4}, garage_mappings={'Missing': 0, 'NA': 0,
'Unf': 1, 'RFn': 2, 'Fin': 3}, finish_mappings={'Missing': 0, 'NA': 0, 'Unf': 1, 'LwQ': 2, 'Rec': 3, 'BLQ': 4, 'ALQ':
5, 'GLQ': 6}))
```

```
data = {
    'MSSubClass': 20,
    'MSZoning': 'RL',
    'LotArea': 7500,
    'LotShape': 'Reg',
    'LandContour': 'Lvl',
    'LotConfig': 'Inside',
    'Neighborhood': 'NAmes',
    'OverallQual': 6,
    'OverallCond': 5,
    'YearRemodAdd': 2000,
    'RoofStyle': 'Gable',
    'Exterior1st': 'VinylSd',
    'ExterQual': 'TA',
    'Foundation': 'PConc',
    'BsmtQual': 'Ex',
    'BsmtExposure': 'Gd',
    'BsmtFinType1': 'GLQ',
    'HeatingQC': 'Ex',
    'CentralAir': 'Y',
    'FirstFlrSF': 856,
    'SecondFlrSF': 854,
    'GrLivArea': 1710,
    'BsmtFullBath': 1,
    'FullBath': 2,
    'HalfBath': 1,
    'KitchenQual': 'TA',
    'TotRmsAbvGrd': 6,
    'Functional': 'Typ',
    'Fireplaces': 1,
    'FireplaceQu': 'Gd',
    'GarageFinish': 'RFn',
```

```
    'GarageCars': 2,
    'PavedDrive': 'Y',
    'WoodDeckSF': 0,
    'ScreenPorch': 0,
    'SaleCondition': 'Normal',
    'LotFrontage':12,
    'GarageArea':14,
    'YrSold':2020
  }


df = pd.DataFrame([data])
df.to_csv('test2.csv')


predict.validate_inputs(input_data=df)

(  MSSubClass MSZoning  LotFrontage LotShape LandContour LotConfig  \
 0         20       RL           12      Reg         Lvl    Inside

   Neighborhood  OverallQual  OverallCond  YearRemodAdd  ... Fireplaces  \
 0        NAmes            6            5          2000  ...          1

   FireplaceQu GarageFinish GarageCars GarageArea PavedDrive WoodDeckSF  \
 0          Gd          RFn          2         14          Y          0

   ScreenPorch SaleCondition  YrSold
 0           0        Normal    2020

 [1 rows x 37 columns],
 None)


predict.make_prediction(input_data=df)

---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
<ipython-input-291-8173e6a8b672> in <cell line: 1>()
----> 1 predict.make_prediction(input_data=df)

 5 frames
/usr/local/lib/python3.10/dist-packages/sklearn/base.py in _check_feature_names(self, X, reset)
    479                 )
    480
--> 481             raise ValueError(message)
    482
    483     def _validate_data(

ValueError: The feature names should match those that were passed during fit.
Feature names unseen at fit time:
- FirstFlrSF
- GarageArea
- LotFrontage
- SecondFlrSF
- YrSold
Feature names seen at fit time, yet now missing:
- 1stFlrSF
- 2ndFlrSF
- FullBath
- LotArea

SEARCH STACK OVERFLOW
```