# **Applied Data Science With Python**

Course-End Project Problem Statement



# **Course-End Project: Feature Engineering**

#### **Project Statement:**

While searching for the dream house, the buyer looks at various factors, not just at the height of the basement ceiling or the proximity to an east-west railroad. Using the dataset, find the factors that influence price negotiations while buying a house.

There are 79 explanatory variables describing every aspect of residential homes in Ames, Iowa.

### **Dataset Description:**

<u>Variable</u>	<u>Description</u>
SalePrice	The property's sale price is in dollars. This is
	the target variable that you're trying to
	predict.
MSSubClass	The building class
MSZoning	The general zoning classification
LotFrontage	Linear feet of street connected to property
LotArea	Lot size in square feet
Street	Type of road access
Alley	Type of alley access
LotShape	General shape of property
LandContour	Flatness of the property
Utilities	Type of utilities available
LotConfig	Lot configuration
LandSlope	Slope of property
Neighborhood	Physical locations within Ames city limits
Condition1	Proximity to main road or railroad
Condition2	Proximity to main road or railroad (if a
	second is present)
BldgType	Type of dwelling
HouseStyle	Style of dwelling
OverallQual	Overall material and finish quality

YearRemidAdd         Remodel date           RoofStyle         Type of roof           RoofMatl         Roof material           Exterior1st         Exterior covering on house           Exterior2nd         Exterior covering on house (if more than one material)           MasVnrType         Masonry veneer type           MasVnrArea         Masonry veneer area in square feet           ExterOual         Exterior material quality           ExterCond         Present condition of the material on the exterior           Foundation         Type of foundation           BsmtQual         Height of the basement           BsmtCond         General condition of the basement           BsmtExposure         Walkout or garden level basement walls           BsmtFinType1         Quality of the basement finished area           BsmtFinSF1         Type 1 finished square feet           BsmtFinSF2         Quality of second finished area (if present)           BsmtFinSF2         Type 2 finished square feet           BsmtUnfSF         Unfinished square feet of basement area           Total square feet of basement area         Total square feet of basement area           Heating         Type of heating           Heating quality and condition         Central air conditioning           Electrical	OverallCond	Overall condition rating
RoofStyle RoofMatl Exterior1st Exterior covering on house Exterior2nd Exterior covering on house (if more than one material) MasVnrType MasVnrArea Masonry veneer type MasVnrArea Exterior material quality ExterCond Exterior Foundation Foundation BsmtQual BsmtExposure BsmtExposure BsmtFinType1 BsmtFinType2 BsmtFinSF1 BsmtFinSF2 BsmtFinSF2 BsmtFinSF2 BsmtFinSF3 Type 1 finished square feet BsmtCond Singular feet BsmtGod General condition of the basement walls BrinSF1 BrinSF3 BrinSF4 BrinSF5 BrinSF5 BrinSF5 BrinSF5 BrinSF5 BrinSF6 BrinSF6 BrinSF7 BrinSF7 BrinSF8 BrinSF8 BrinSF9 Br	YearBuilt	Original construction date
RoofMatl Exterior1st Exterior covering on house Exterior covering on house (if more than one material)  MasVnrType MasVnrArea MasOnry veneer type MasVnrArea Exterior covering on house (if more than one material)  MasVnrType MasOnry veneer type MasVnrArea ExterQual Exterior material quality  ExterCond Present condition of the material on the exterior Foundation Type of foundation BsmtQual Height of the basement BsmtCond General condition of the basement BsmtExposure Walkout or garden level basement walls BsmtFinType1 Quality of the basement finished area BsmtFinType2 Quality of second finished square feet BsmtFinSF1 Type 1 finished square feet BsmtFinSF2 BsmtJnfSF Type 2 finished square feet TotalBsmtSF Unfinished square feet of basement area TotalBsmtSF Total square feet of basement area TotalSgmtSF Total square feet of basement area Heating Type of heating HeatingQC Heating quality and condition CentralAir Central air conditioning Electrical Electrical Electrical system IstFISF Second floor square feet Low QualFinSF Low quality finished square feet (all floors) Above grade (ground) living area square feet BsmtFullBath Basement full bathrooms BsmtHalfBath Basement full bathrooms Full Bath Full bathrooms above grade HalfBath Half bathrooms above basement	YearRemodAdd	Remodel date
Exterior1st Exterior covering on house  Exterior2nd Exterior covering on house (if more than one material)  MasVnrType MasVnrArea Masonry veneer type  MasVnrArea Exterior material quality  ExterCond ExterCond Present condition of the material on the exterior  Foundation Type of foundation  BsmtQual Height of the basement  BsmtCond General condition of the basement  BsmtExposure Walkout or garden level basement walls  BsmtFinType1 Quality of the basement finished area  BsmtFinSF1 Type 1 finished square feet  BsmtFinSF2 Quality of second finished area (if present)  BsmtFinSF2 Type 2 finished square feet  BsmtUnfSF Unfinished square feet of basement area  TotalBsmtSF Total square feet of basement area  Heating Type of heating  Heating QC Heating quality and condition  CentralAir Central air conditioning  Electrical Electrical Electrical system  IstFIrSF Second floor square feet  LowQualFinSF Low quality finished square feet (all floors)  Above grade (ground) living area square feet  BsmtFullBath Basement half bathrooms  BsmtHalfBath Full bathrooms above grade  HalfBath Half bathrooms above grade  Number of bedrooms above basement	RoofStyle	Type of roof
Exterior covering on house (if more than one material)  MasVnrType MasVnrArea Masonry veneer type MasVnrArea ExterQual Exterior material quality ExterCond Present condition of the material on the exterior Foundation BsmtQual BsmtCond General condition of the basement BsmtExposure Walkout or garden level basement walls BsmtFinType1 Quality of the basement finished area BsmtFinType2 Quality of second finished area (if present) BsmtFinSF2 BsmtFinSF2 BsmtJnfSF Unfinished square feet Unfinished square feet of basement area TotalBsmtSF Total square feet of basement area Heating Heating HeatingQC Heating quality and condition CentralAir Central air conditioning Electrical IstElrSF Second floor square feet LowQualFinSF Low quality finished square feet (all floors) Above grade (ground) living area square feet BsmtFullBath Basement half bathrooms BsmtPalfBath Full bathrooms above grade Number of bedrooms above basement	RoofMatl	Roof material
masVnrType MasVnrType MasVnrArea MasOnry veneer area in square feet ExterQual ExterCond ExterCond ExterCond Foundation BsmtQual BsmtCond BsmtExposure BsmtExposure BsmtFinType1 BsmtFinType2 BsmtFinType2 BsmtFinSF2 BsmtSF5 BsmtSF5 BsmtSF5 BsmtSF6 BsmtSF6 BsmtSF7 BsmtSF7 BsmtSF7 BsmtSF7 BsmtSF8 BsmtInType1 BsmtSF8 BsmtInType2 BsmtInType2 BsmtInType3 BsmtInType3 BsmtInType4 BsmtInType5 BsmtInType6 BsmtInType7 BsmtInType7 BsmtInType8 BsmtInType8 BsmtInType8 BsmtInType9 BymtInType9 BymtI	Exterior1st	Exterior covering on house
MasVnrTypeMasonry veneer typeMasVnrAreaMasonry veneer area in square feetExterQualExterior material qualityExterCondPresent condition of the material on the exteriorFoundationType of foundationBsmtQualHeight of the basementBsmtCondGeneral condition of the basementBsmtExposureWalkout or garden level basement wallsBsmtFinType1Quality of the basement finished areaBsmtFinSF1Type 1 finished square feetBsmtFinType2Quality of second finished area (if present)BsmtFinSF2Type 2 finished square feetBsmtUnfSFUnfinished square feet of basement areaTotal Square feet of basement areaTotal square feet of basement areaHeatingType of heatingHeatingQCHeating quality and conditionCentralAirCentral air conditioningElectricalElectrical system1stFIrSFFirst Floor square feetLow qualFinSFSecond floor square feetLow qualFinSFLow quality finished square feet (all floors)GrLivAreaAbove grade (ground) living area square feetBsmtFullBathBasement full bathroomsBsmtFullBathBasement half bathroomsBsmtHalfBathFull bathrooms above gradeHalfBathHalf bathrooms above basement	Exterior2nd	
MasVnrAreaMasonry veneer area in square feetExterQualExterior material qualityExterCondPresent condition of the material on the exteriorFoundationType of foundationBsmtQualHeight of the basementBsmtExposureWalkout or garden level basement wallsBsmtExposureQuality of the basement finished areaBsmtFinType1Quality of the basement finished areaBsmtFinSF1Type 1 finished square feetBsmtFinSF2Quality of second finished area (if present)BsmtFinSF2Type 2 finished square feetBsmtUnfSFUnfinished square feet of basement areaTotalBsmtSFTotal square feet of basement areaHeatingType of heatingHeatingQCHeating quality and conditionCentralAirCentral air conditioningElectricalElectrical system1stFIrSFFirst Floor square feet2ndFIrSFSecond floor square feetLowQualFinSFLow quality finished square feet (all floors)GrLivAreaAbove grade (ground) living area square feetBsmtFullBathBasement full bathroomsBsmtHalfBathBasement half bathroomsFull bathrooms above gradeHalfBathHalf bathrooms above basement	MasVnrTvne	,
ExterQual Exterior material quality  Present condition of the material on the exterior  Foundation Type of foundation  BsmtQual Height of the basement  BsmtCond General condition of the basement  BsmtExposure Walkout or garden level basement walls  BsmtFinType1 Quality of the basement finished area  BsmtFinSF1 Type 1 finished square feet  BsmtFinType2 Quality of second finished area (if present)  BsmtFinSF2 Type 2 finished square feet  BsmtUnfSF Unfinished square feet of basement area  TotalBsmtSF Total square feet of basement area  TotalBsmtSF Total square feet of basement area  Heating Type of heating  HeatingQC Heating quality and condition  CentralAir Central air conditioning  Electrical Electrical system  1stFIrSF First Floor square feet  LowQualFinSF Second floor square feet  LowQualFinSF Low quality finished square feet (all floors)  GrLivArea Above grade (ground) living area square feet  BsmtFullBath Basement full bathrooms  BsmtHalfBath Basement half bathrooms  Full bathrooms above grade  HalfBath Half bathrooms above basement		
ExterCond  Present condition of the material on the exterior  Type of foundation  BsmtQual  BsmtCond  BsmtExposure  BsmtFinType1  BsmtFinSF1  BsmtFinType2  BymtFinSF2  BymtFinSF2  BymtFinSF3  BymtFinSF3  BymtFinSF4  BymtFinSF5  BymtFinSF5  BymtFinSF5  BymtFinSF6  BymtFinSF6  BymtFinSF7  BymtFinSF7  BymtFinSF8  BymtFinSF8  BymtFinSF8  BymtFinSF8  BymtFinSF9  BymtFinSF9  BymtFinSF9  BymtFinSF9  BymtFinSF9  BymtFinSF9  BymtJnished square feet  BymtJnished square feet of basement area  Total square feet of basement area  Total square feet of basement area  Total square feet of basement area  Type of heating  Heating  Heating  Central air conditioning  Electrical  Electrical system  1stFIrSF  First Floor square feet  2ndFIrSF  Second floor square feet  LowQualFinSF  Low quality finished square feet (all floors)  Above grade (ground) living area square feet  BsmtFullBath  Basement full bathrooms  BsmtHalfBath  Basement half bathrooms  Full bathrooms above grade  Number of bedrooms above basement		
ExterCond Foundation Foundation Type of foundation  BsmtQual BsmtCond General condition of the basement  BsmtExposure Walkout or garden level basement walls  BsmtFinType1 Quality of the basement finished area  BsmtFinSF1 Type 1 finished square feet  BsmtFinSF2 Quality of second finished area (if present)  BsmtFinSF2 Type 2 finished square feet  BsmtUnfSF Unfinished square feet of basement area  TotalBsmtSF Total square feet of basement area  Totalsquare feet of basement area  Heating HeatingQC Heating quality and condition  CentralAir Central air conditioning  Electrical Electrical system  1stFIrSF First Floor square feet  LowQualFinSF Low quality finished square feet (all floors)  Above grade (ground) living area square feet  BsmtFullBath Basement full bathrooms  BsmtHalfBath Full bathrooms above grade  Half bathrooms above grade  Number of bedrooms above basement	ExterQual	
BsmtQual Height of the basement  BsmtCond General condition of the basement  BsmtExposure Walkout or garden level basement walls  BsmtFinType1 Quality of the basement finished area  BsmtFinSF1 Type 1 finished square feet  BsmtFinType2 Quality of second finished area (if present)  BsmtFinSF2 Type 2 finished square feet  BsmtUnfSF Unfinished square feet of basement area  TotalBsmtSF Total square feet of basement area  TotalBsmtSF Total square feet of basement area  Heating Type of heating  HeatingQC Heating quality and condition  CentralAir Central air conditioning  Electrical Electrical system  1stFrSF First Floor square feet  2ndFlrSF Second floor square feet  LowQualFinSF Low quality finished square feet (all floors)  GrLivArea feet  BsmtFullBath Basement full bathrooms  BsmtHalfBath Basement half bathrooms  Full bathrooms above grade  HalfBath Half bathrooms above grade  Number of bedrooms above basement	ExterCond	
BsmtQual Height of the basement  BsmtCond General condition of the basement  BsmtExposure Walkout or garden level basement walls  BsmtFinType1 Quality of the basement finished area  BsmtFinSF1 Type 1 finished square feet  BsmtFinType2 Quality of second finished area (if present)  BsmtFinSF2 Type 2 finished square feet  BsmtUnfSF Unfinished square feet of basement area  TotalBsmtSF Total square feet of basement area  TotalBsmtSF Total square feet of basement area  Heating Type of heating  HeatingQC Heating quality and condition  CentralAir Central air conditioning  Electrical Electrical system  1stFrSF First Floor square feet  2ndFlrSF Second floor square feet  LowQualFinSF Low quality finished square feet (all floors)  GrLivArea feet  BsmtFullBath Basement full bathrooms  BsmtHalfBath Basement half bathrooms  Full bathrooms above grade  HalfBath Half bathrooms above grade  Number of bedrooms above basement	Foundation	Type of foundation
BsmtCond BsmtExposure Walkout or garden level basement Walkout or garden level basement walls BsmtFinType1 Quality of the basement finished area BsmtFinSF1 Type 1 finished square feet  BsmtFinSF2 Type 2 finished square feet  BsmtUnfSF Unfinished square feet of basement area TotalBsmtSF Total square feet of basement area Theating Heating Heating Heating CentralAir Electrical Electrical StFIrSF First Floor square feet LowQualFinSF Low quality finished square feet (all floors) Above grade (ground) living area square feet  BsmtFullBath Basement full bathrooms Brull bathrooms Full bathrooms above grade HalfBath Half bathrooms above basement  Bedroom  Walkout or garden level basement walls Walkout or garden level basement finished area Unfinished square feet Unfinished square feet Second floor square feet Second floor square feet Low quality finished square feet (all floors) Above grade (ground) living area square feet BsmtFullBath Basement full bathrooms Brull bathrooms Brull bathrooms above grade Half bathrooms above basement	BsmtQual	
BsmtExposure Walkout or garden level basement walls BsmtFinType1 Quality of the basement finished area BsmtFinSF1 Type 1 finished square feet BsmtFinType2 Quality of second finished area (if present) BsmtFinSF2 Type 2 finished square feet BsmtUnfSF Unfinished square feet of basement area TotalBsmtSF Total square feet of basement area Heating Type of heating HeatingQC Heating quality and condition CentralAir Central air conditioning Electrical Electrical system 1stFIrSF First Floor square feet LowQualFinSF Second floor square feet LowQualFinSF Low quality finished square feet (all floors) Above grade (ground) living area square feet BsmtFullBath Basement full bathrooms BsmtHalfBath Basement half bathrooms FullBath Half bathrooms above grade HalfBath Half bathrooms above basement		
BsmtFinType1 Quality of the basement finished area BsmtFinSF1 Type 1 finished square feet BsmtFinType2 Quality of second finished area (if present) BsmtFinSF2 Type 2 finished square feet BsmtUnfSF Unfinished square feet of basement area TotalBsmtSF Total square feet of basement area Heating Type of heating HeatingQC Heating quality and condition CentralAir Central air conditioning Electrical Electrical system 1stFIrSF First Floor square feet 2ndFIrSF Second floor square feet LowQualFinSF Low quality finished square feet (all floors) Above grade (ground) living area square feet BsmtFullBath Basement full bathrooms BsmtHalfBath Basement half bathrooms FullBath Half bathrooms above grade HalfBath Half bathrooms above basement		
BsmtFinSF1 Type 1 finished square feet  BsmtFinType2 Quality of second finished area (if present)  BsmtFinSF2 Type 2 finished square feet  BsmtUnfSF Unfinished square feet of basement area  TotalBsmtSF Total square feet of basement area  Heating Type of heating  HeatingQC Heating quality and condition  CentralAir Central air conditioning  Electrical Electrical system  1stFIrSF First Floor square feet  2ndFIrSF Second floor square feet  LowQualFinSF Low quality finished square feet (all floors)  GrLivArea Basement full bathrooms  BsmtHalfBath Basement half bathrooms  Full bathrooms above grade  HalfBath Half bathrooms above grade  Number of bedrooms above basement	·	
BsmtFinType2 Quality of second finished area (if present) BsmtFinSF2 Type 2 finished square feet BsmtUnfSF Unfinished square feet of basement area TotalBsmtSF Total square feet of basement area Heating Type of heating HeatingQC Heating quality and condition CentralAir Central air conditioning Electrical Electrical system 1stFIrSF First Floor square feet 2ndFIrSF Second floor square feet LowQualFinSF Low quality finished square feet (all floors) GrLivArea BsmtFullBath Basement full bathrooms BsmtHalfBath Basement half bathrooms FullBath Full bathrooms above grade HalfBath Half bathrooms above grade Number of bedrooms above basement		
BsmtFinSF2 Type 2 finished square feet BsmtUnfSF Unfinished square feet of basement area TotalBsmtSF Total square feet of basement area Heating Type of heating HeatingQC Heating quality and condition CentralAir Central air conditioning Electrical Electrical system 1stFlrSF First Floor square feet 2ndFlrSF Second floor square feet LowQualFinSF Low quality finished square feet (all floors) GrLivArea Basement full bathrooms BsmtFullBath Basement half bathrooms FullBath Full bathrooms above grade HalfBath Half bathrooms above grade Number of bedrooms above basement	BsmtFinType2	
TotalBsmtSF Heating Type of heating HeatingQC Heating quality and condition CentralAir Central air conditioning Electrical Electrical Electrical system 1stFIrSF Second floor square feet LowQualFinSF Low quality finished square feet (all floors) Above grade (ground) living area square feet BsmtFullBath Basement full bathrooms BsmtHalfBath Full bathrooms above grade HalfBath Half bathrooms above grade Number of bedrooms above basement		Type 2 finished square feet
TotalBsmtSF Heating Type of heating HeatingQC Heating quality and condition CentralAir Central air conditioning Electrical Electrical Electrical system First Floor square feet  2ndFlrSF Second floor square feet LowQualFinSF Low quality finished square feet (all floors) Above grade (ground) living area square feet BsmtFullBath Basement full bathrooms BsmtHalfBath Full bathrooms above grade HalfBath Half bathrooms above grade Number of bedrooms above basement	BsmtUnfSF	Unfinished square feet of basement area
Heating QC Heating quality and condition  CentralAir Central air conditioning  Electrical Electrical system  1stFIrSF First Floor square feet  2ndFIrSF Second floor square feet  LowQualFinSF Low quality finished square feet (all floors)  GrLivArea Above grade (ground) living area square feet  BsmtFullBath Basement full bathrooms  BsmtHalfBath Basement half bathrooms  FullBath Full bathrooms above grade  HalfBath Half bathrooms above grade  Number of bedrooms above basement	TotalBsmtSF	Total square feet of basement area
CentralAirCentral air conditioningElectricalElectrical system1stFlrSFFirst Floor square feet2ndFlrSFSecond floor square feetLowQualFinSFLow quality finished square feet (all floors)GrLivAreaAbove grade (ground) living area square feetBsmtFullBathBasement full bathroomsBsmtHalfBathBasement half bathroomsFullBathFull bathrooms above gradeHalfBathHalf bathrooms above basement	Heating	
CentralAirCentral air conditioningElectricalElectrical system1stFlrSFFirst Floor square feet2ndFlrSFSecond floor square feetLowQualFinSFLow quality finished square feet (all floors)GrLivAreaAbove grade (ground) living area square feetBsmtFullBathBasement full bathroomsBsmtHalfBathBasement half bathroomsFullBathFull bathrooms above gradeHalfBathHalf bathrooms above basement	HeatingQC	
1stFlrSFFirst Floor square feet2ndFlrSFSecond floor square feetLowQualFinSFLow quality finished square feet (all floors)GrLivAreaAbove grade (ground) living area square feetBsmtFullBathBasement full bathroomsBsmtHalfBathBasement half bathroomsFullBathFull bathrooms above gradeHalfBathHalf bathrooms above gradeNumber of bedrooms above basement	CentralAir	
2ndFlrSFSecond floor square feetLowQualFinSFLow quality finished square feet (all floors)GrLivAreaAbove grade (ground) living area square feetBsmtFullBathBasement full bathroomsBsmtHalfBathBasement half bathroomsFullBathFull bathrooms above gradeHalfBathHalf bathrooms above gradeNumber of bedrooms above basement	Electrical	Electrical system
Low Quality finished square feet (all floors)  Above grade (ground) living area square feet  BsmtFullBath  Basement full bathrooms  BsmtHalfBath  Full bathrooms above grade  HalfBath  Half bathrooms above grade  Number of bedrooms above basement	1stFlrSF	First Floor square feet
GrLivArea  Above grade (ground) living area square feet  BsmtFullBath  Basement full bathrooms  BsmtHalfBath  Full bathrooms above grade  HalfBath  Half bathrooms above grade  Number of bedrooms above basement	2ndFlrSF	Second floor square feet
BsmtFullBath BsmtHalfBath Basement full bathrooms BsmtHalfBath Basement half bathrooms FullBath Full bathrooms above grade HalfBath Half bathrooms above grade Number of bedrooms above basement	LowQualFinSF	Low quality finished square feet (all floors)
BsmtFullBath Basement full bathrooms BsmtHalfBath Basement half bathrooms FullBath Full bathrooms above grade HalfBath Half bathrooms above grade Number of bedrooms above basement	GrLivArea	
BsmtHalfBath  FullBath  Full bathrooms above grade  HalfBath  Half bathrooms above grade  Number of bedrooms above basement	BsmtFullBath	
FullBath Full bathrooms above grade HalfBath Half bathrooms above grade Number of bedrooms above basement		
HalfBath Half bathrooms above grade  Number of bedrooms above basement		
Number of bedrooms above basement		
[ [X. W X.1		Number of bedrooms above basement

Kitchen	Number of kitchens
KitchenQual	Kitchen quality
TotRmsAbvGrd	Total rooms above grade (does not include
TOTRITISADVGI U	bathrooms)
Functional	Home functionality rating
Fireplaces	Number of fireplaces
FireplaceQu	Fireplace quality
GarageType	Garage location
GarageYrBlt	Year garage was built
GarageFinish	Interior finish of the garage
GarageCars	Size of the garage in car capacity
GarageArea	Size of the garage in square feet
GarageQual	Garage quality
GarageCond	Garage condition
PavedDrive	Paved driveway
WoodDeckSF	Wood deck area in square feet
OpenPorchSF	Open porch area in square feet
EnclosedPorch	Enclosed porch area in square feet
3SsnPorch	Three season porch area in square feet
ScreenPorch	Screen porch area in square feet
PoolArea	Pool area in square feet
PoolQC	Pool quality
Fence	Fence quality
MiscFeature	Miscellaneous feature not covered in other
Miscreature	categories
MiscVal	\$Value of miscellaneous feature
MoSold	Month Sold
YrSold	Year Sold
SaleType	Type of sale
SaleCondition	Condition of sale

## Note:

1) Download the "PEP1.csv" using the link given in the Feature Engineering project problem statement

2) For a detailed description of the dataset, you can download and refer to data\_description.txt using the link given in the Feature Engineering project problem statement

### Perform the following steps:

- 1. Understand the dataset:
  - a. Identify the shape of the dataset
  - b. Identify variables with null values
  - c. Identify variables with unique values
- 2. Generate a separate dataset for numerical and categorical variables
- 3. EDA of numerical variables:
  - a. Missing value treatment
  - b. Identify the skewness and distribution
  - c. Identify significant variables using a correlation matrix
  - d. Pair plot for distribution and density
- 4. EDA of categorical variables
  - a. Missing value treatment
  - b. Count plot for bivariate analysis
  - c. Identify significant variables using p-values and Chi-Square values
- 5. Combine all the significant categorical and numerical variables
- 6. Plot box plot for the new dataset to find the variables with outliers

**Note:** The last two points are performed to make the new dataset ready for training and prediction.