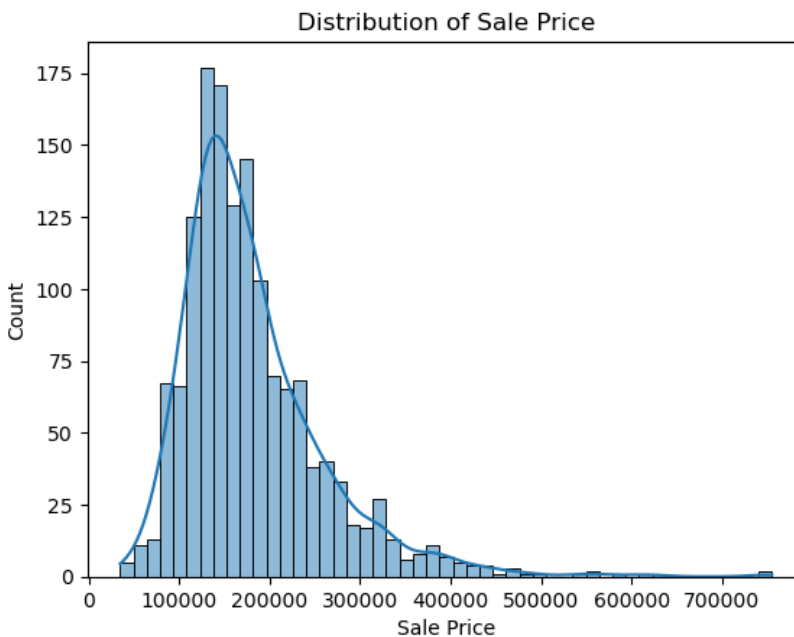# Maryam Gadimova

Assignment_3_Part_2

The exploratory data analysis (EDA) of the House Price Dataset

- The exploratory data analysis (EDA) of the House Price Dataset revealed valuable insights into the dataset's characteristics, distributions, and correlations. The dataset contains information on various features related to houses, with 'SalePrice' serving as the target variable representing the sale price of the houses.
- The dataset consists of 1460 entries and contains both numerical and categorical features.

	PoolArea	MiscVal	MoSold	YrSold	SalePrice
count	1460.000000	1460.000000	1460.000000	1460.000000	1460.000000
mean	2.758904	43.489041	6.321918	2007.815753	180921.195890
std	40.177307	496.123024	2.703626	1.328095	79442.502883
min	0.000000	0.000000	1.000000	2006.000000	34900.000000
25%	0.000000	0.000000	5.000000	2007.000000	129975.000000
50%	0.000000	0.000000	6.000000	2008.000000	163000.000000
75%	0.000000	0.000000	8.000000	2009.000000	214000.000000
max	738.000000	15500.000000	12.000000	2010.000000	755000.000000

On the summary of data set you can see target value which is 'SalePrice' with the minimum sale price at approximately $34,900 and the maximum at around $755,000. The median sale price is approximately $163,000.

The distribution of 'SalePrice' is right-skewed it means that: Median < Mean.

Distribution of Sale Price

```
print(correlations)
OverallQual      0.790982
GrLivArea        0.708624
GarageCars       0.640409
GarageArea       0.623431
TotalBsmtSF      0.613581
1stFlrSF         0.605852
FullBath         0.560664
TotRmsAbvGrd     0.533723
YearBuilt        0.522897
YearRemodAdd     0.507101
GarageYrBlt      0.486362
MasVnrArea       0.477493
Fireplaces       0.466929
BsmtFinSF1       0.386420
LotFrontage      0.351799
WoodDeckSF       0.324413
2ndFlrSF         0.319334
OpenPorchSF      0.315856
HalfBath         0.284108
LotArea          0.263843
BsmtFullBath     0.227122
BsmtUnfSF        0.214479
BedroomAbvGr     0.168213
ScreenPorch      0.111447
PoolArea         0.092404
MoSold           0.046432
3SsnPorch        0.044584
BsmtFinSF2      -0.011378
BsmtHalfBath    -0.016844
MiscVal         -0.021190
Id              -0.021917
LowQualFinSF    -0.025606
YrSold          -0.028923
OverallCond     -0.077856
MSSubClass      -0.084284
EnclosedPorch   -0.128578
KitchenAbvGr    -0.135907
Name: SalePrice, dtype: float64
```

Correlation shows the relationships between numerical variables.

You can see the 'OverallQual' is the most positively correlated feature with 'SalePrice'

On the other hand, you can see the 'KitchenAbvGr' has the most negative correlation.

The bar chart illustrates the relationship between 'ExterQual' (Exterior Quality) and 'MasVnrType' (Masonry Veneer Type).

- The chart shows the distribution of 'MasVnrType' categories ('Ex', 'Fa', 'Gd', 'TA') for each level of 'ExterQual.'
- It is a non-stacked bar chart, which allows for a clear comparison of 'MasVnrType' categories within each 'ExterQual' level.

In summary, this chart provides insights into the relationships between exterior quality and masonry veneer types within the dataset. It helps in understanding which types of masonry veneer are commonly found in houses with different exterior quality levels. These insights can be valuable for architectural and construction decisions, especially in assessing the compatibility of masonry veneer with various exterior quality standards. Further analysis or domain-specific expertise may be required to make informed decisions based on these observations.