

# Diamonds project

Baraa Hammoud  
Khaled Mohamed  
Fidaa Ghazlan



- EDA
- Preprocessing
- Training model

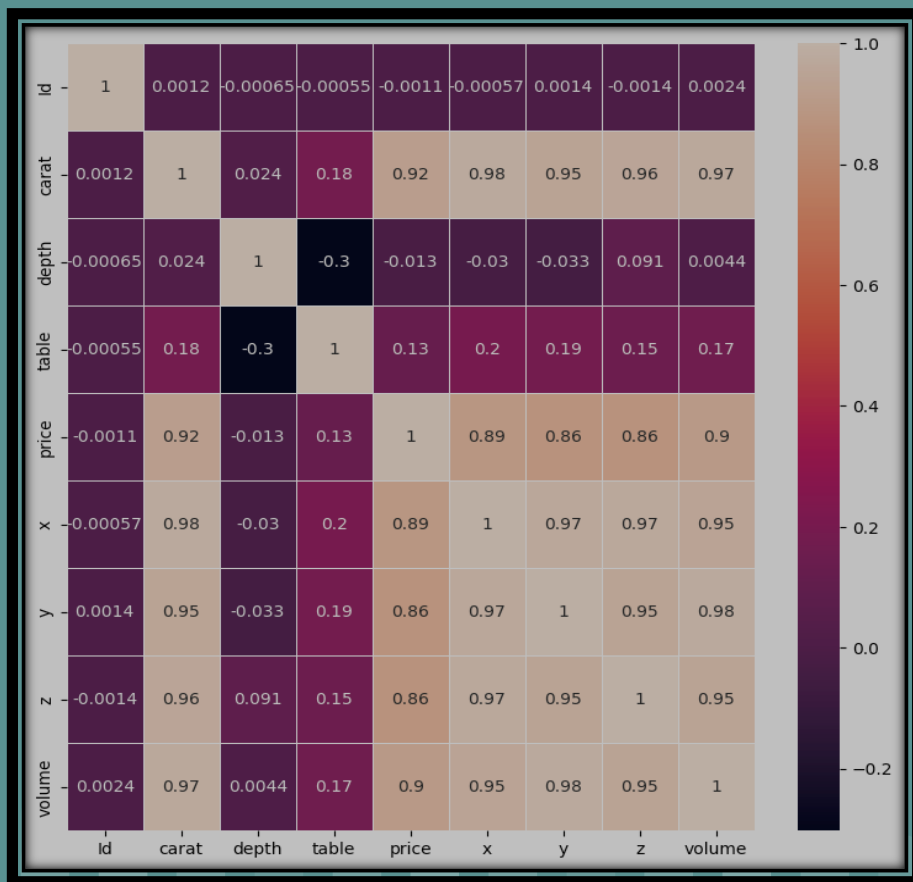
EDA:

- working on identifying the potential of each feature
- correlation between features
- data health check
- statistical info about the data

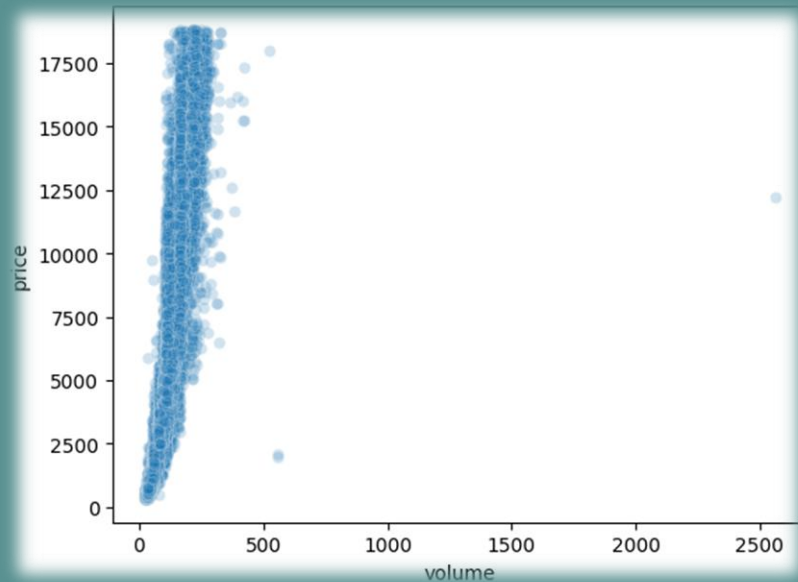
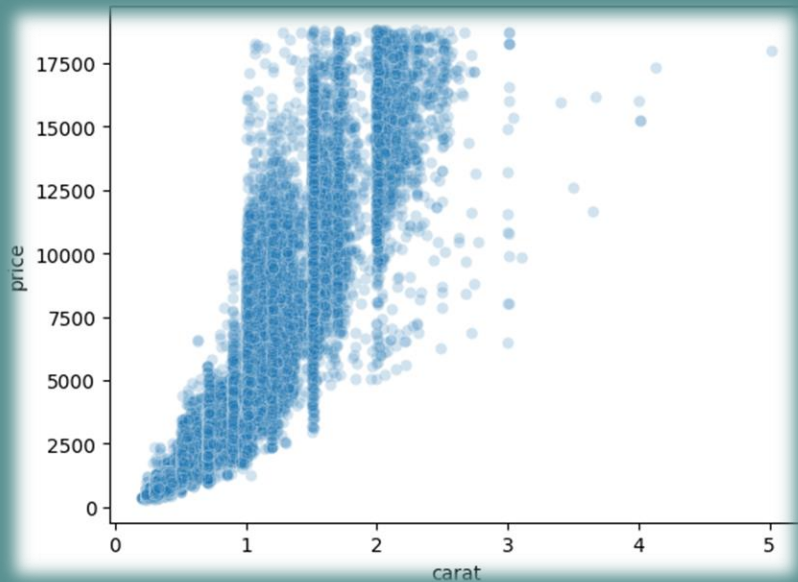


# Correlation chart:

#Correlation after feature selection and deduction

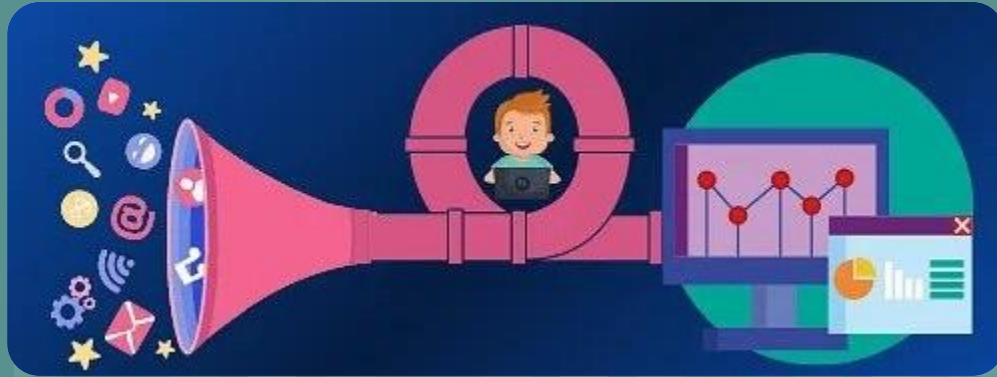


## The most distinguishable features (including added features)

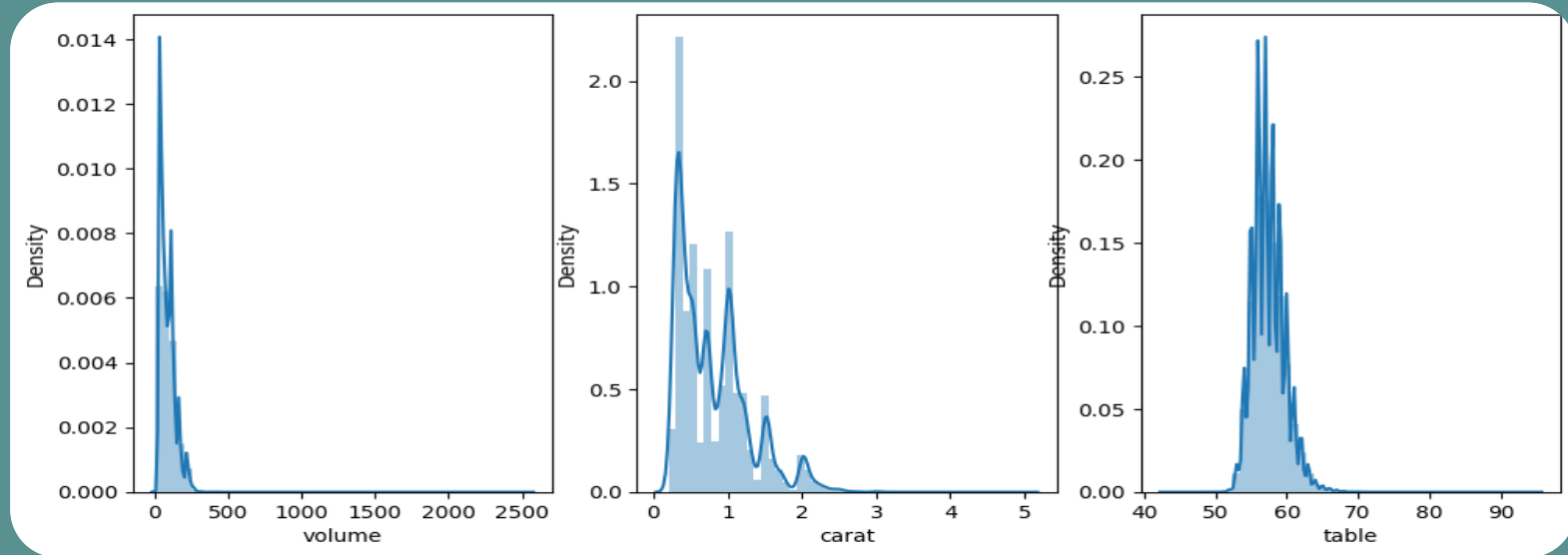


## PREPROCESSING:

- Covert object type to float(one hot encoder)
- Remove outliers(Z Score)
- Normalize the data(Z Score)



# Outliers removal using Z-score method



## Training Phase:

- Linear Regression
- Lasso Regression (regularization)
- Ridge Regression (regularization)
- Random forest tree
- Decision tree





## Evaluation methods:

- Root Mean Square Error
- Cross validation

**Thank you**