Asmaa , Alaa , Saba Team

PREDICTING DIAMOND PRICE

this project part of shai traning club

Before we start

Presentation Parts

- Deep understanding of data
- Data Cleaning
- Data Manipulation
- Data Visualization
- Model Training
- Model Testing In real Data

DATASET PRELIMINARY EXPLANATION

	Unnamed: 0	carat	cut	color	clarity	depth	table	price	x	y	z
0	2	0.21	Premium	E	SI1	59.8	61.0	326	3.89	3.84	2.31
1	4	0.29	Premium	I	VS2	62.4	58.0	334	4.20	4.23	2.63
2	5	0.31	Good	J	SI2	63.3	58.0	335	4.34	4.35	2.75
3	6	0.24	Very Good	J	VVS2	62.8	57.0	336	3.94	3.96	2.48
4	7	0.24	Very Good	I	VVS1	62.3	57.0	336	3.95	3.98	2.47

carat : Piece Weight
cut : Quality
color : Diamond Color
z: depth

calarity :
table :
x: length
y: width

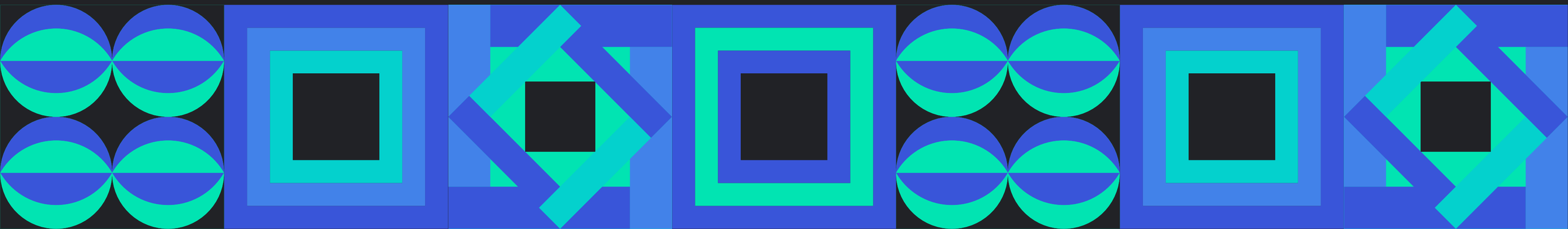
DATA CLEANING

Check Nullable Data

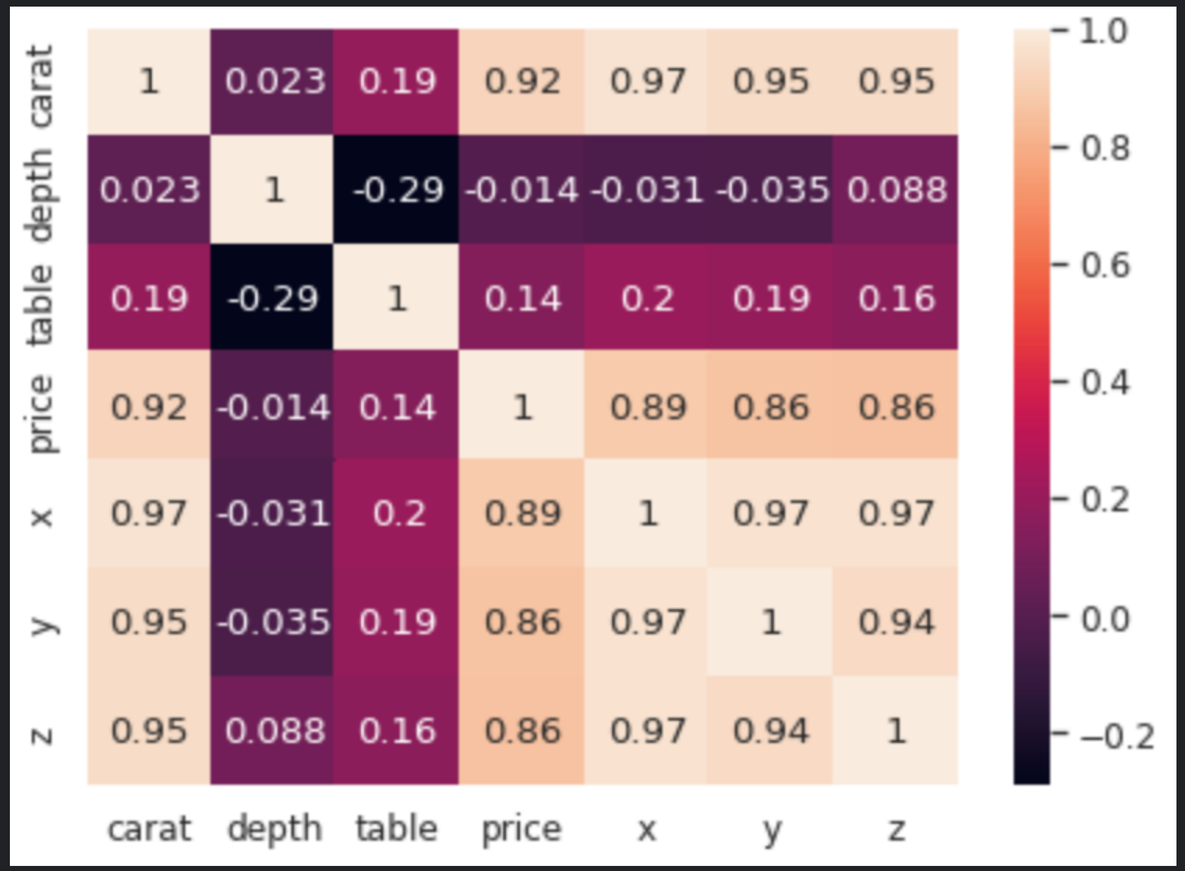
Drop Unwanted Column

check duplicate

Remove Outliers



DATA CLEANING



- This image appear the coorelation between features using heatmap function .

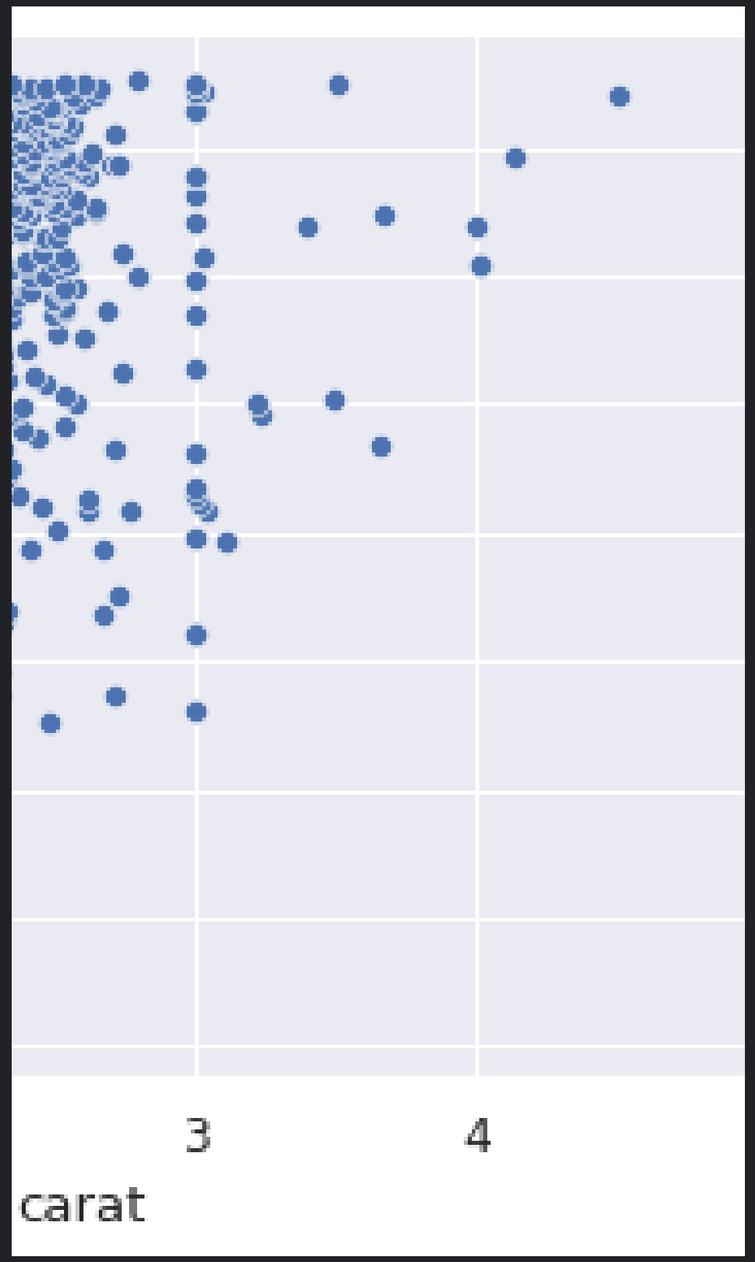
REMOVE OUTLIERS

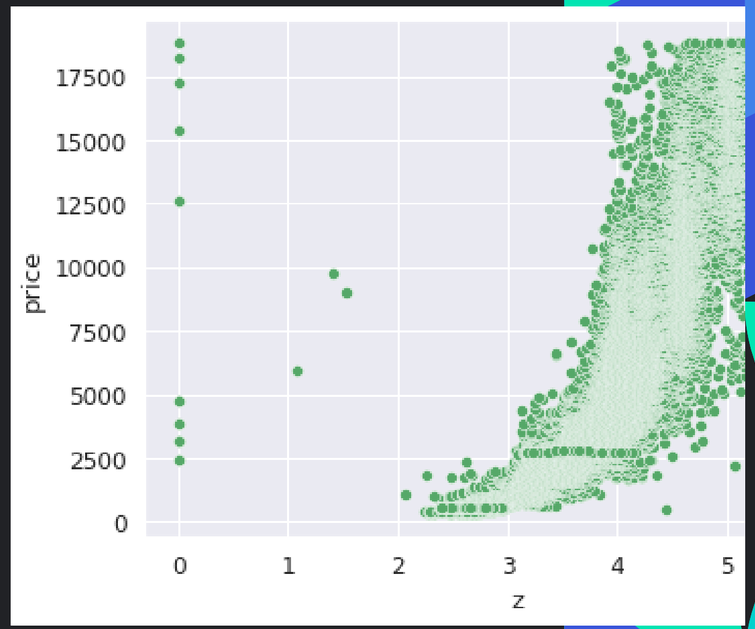
- in Figure 1 remove data > 3 in carat

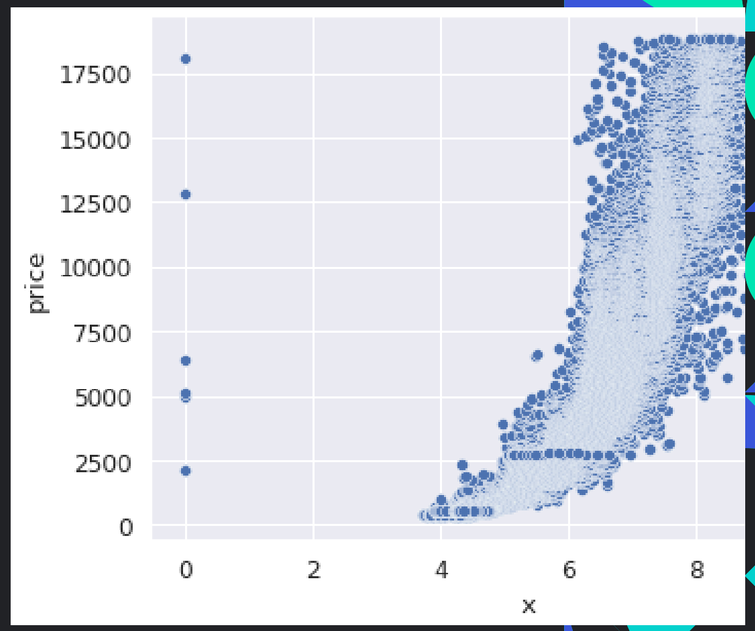
- in Figure 2 remove data <2 in depth

- in Figure 3 remove data <2 in height

- it is important to remove those outliers because the presence of those outliers can mislead the model.





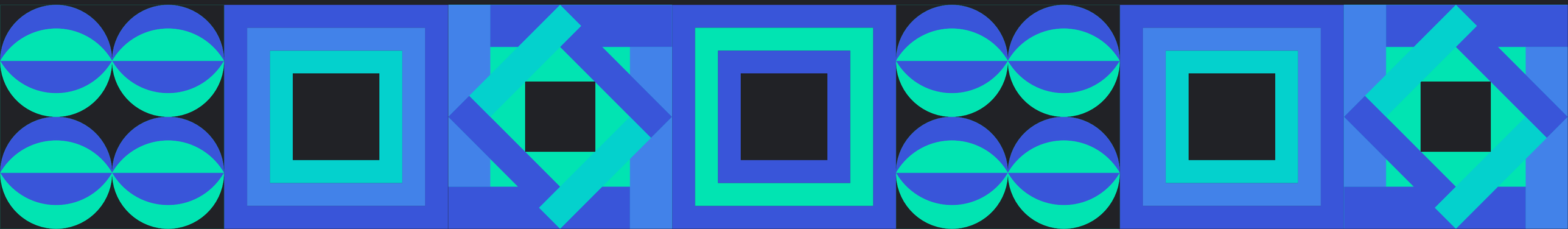


DATA MANIPULATION

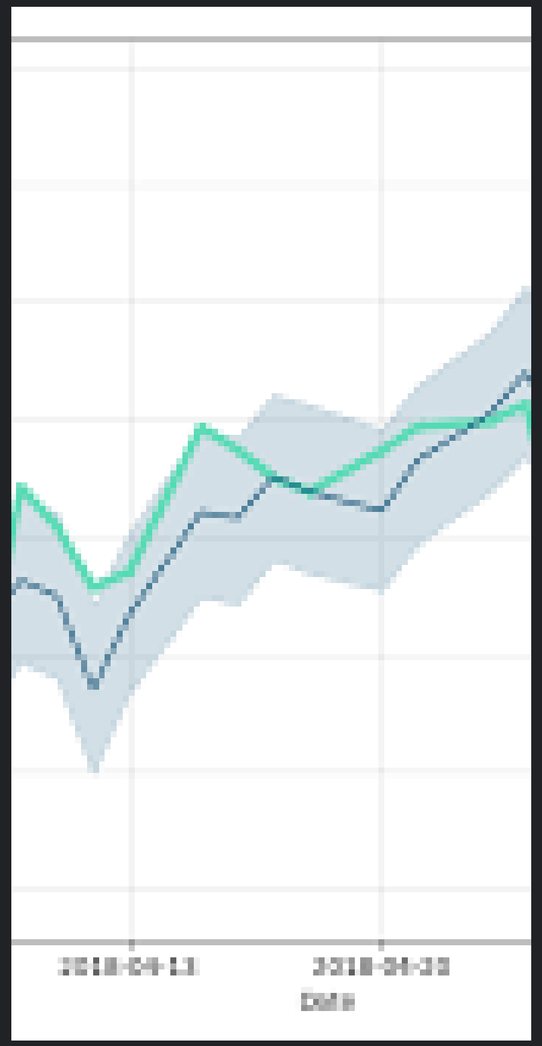
Convert Categorical Data to numerical some model take numerical only

Adding Converted Column to data frame

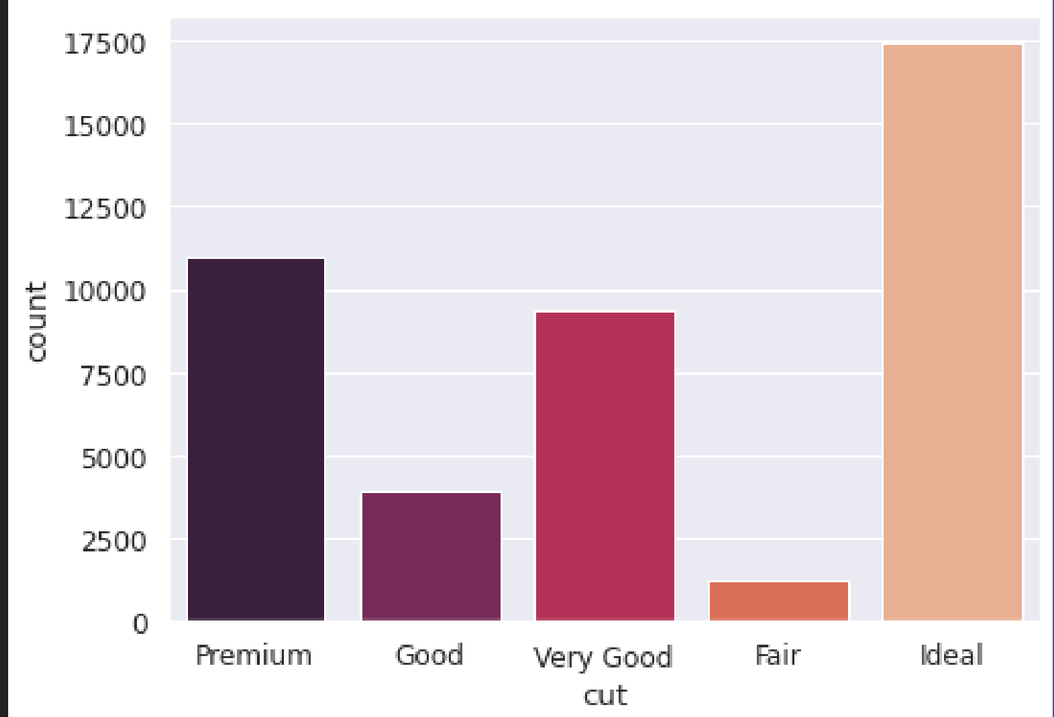
Delete the Categorical Column



DATA VISUALIZATION



CATEGORICAL DATA



POWER PI VISUALIZE

Clear Quality

8

Color

7

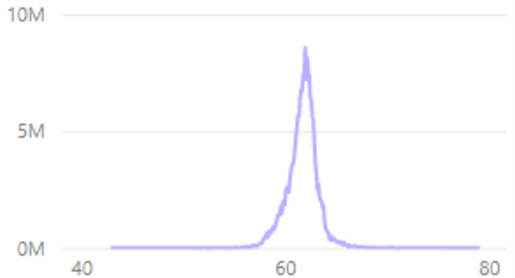
Cut Grades

5

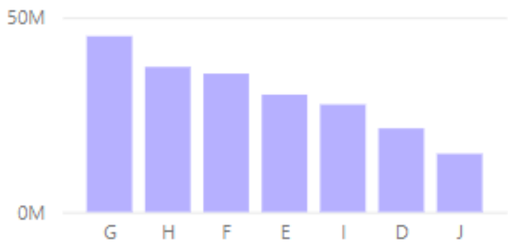
Clarity & Color

price & color

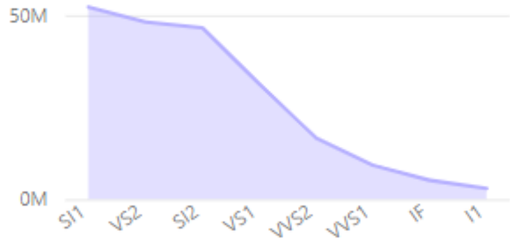
Price & depth



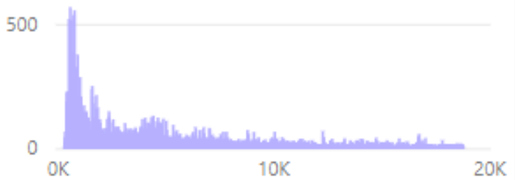
Price & Color



Price & Clarity



Length & Price



Width & Price

Max price & color

D

18693

Max of price

E

18731

Max of price

F

18791

Max of price

G

18818

Max of price

H

18803

Max of price

I

18823

Max of price

J

18710

Max of price

MODEL TRAINING

DECISION TREE

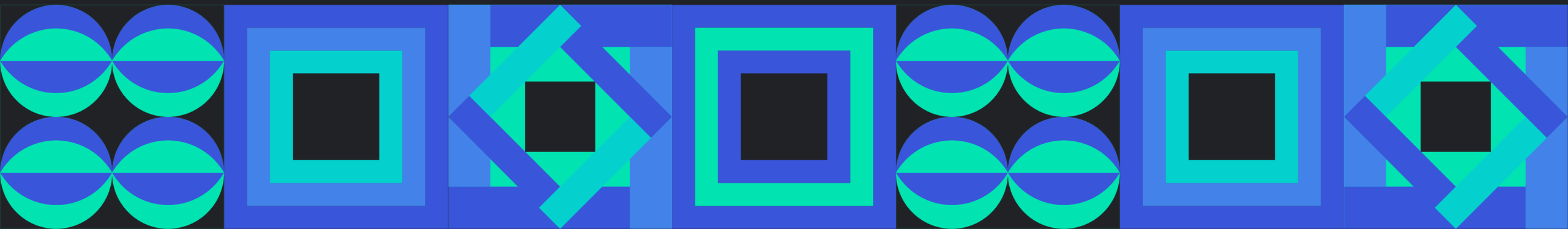
Model train accuracy score is : 0.99

Model test accuracy score is : 0.96

mse = 569830.2647501656

rmse = 754.8710252421705

R2_Score is: 0.9639



MODEL TRAINING

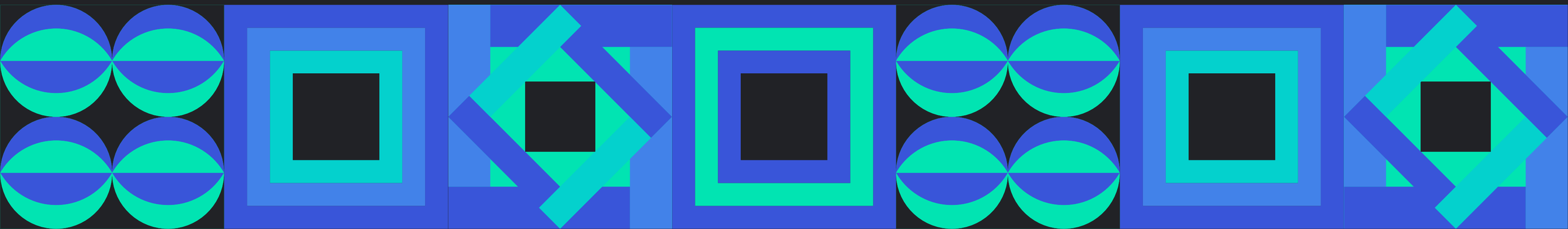
LINEAR REGRESSION

Model train accuracy score is : 0.916089470206155

Model test accuracy score is : 0.9164134961816611

mse = 1320769.2167478143

rmse = 1149.2472391734575



MODEL TRAINING

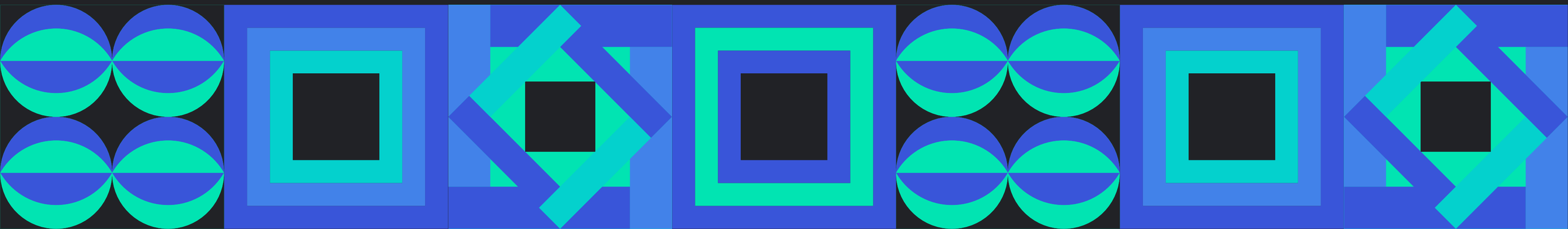
SUPPORT VECTOR MACHINES (SVM)

Model train accuracy score is : 0.510840078518844

Model test accuracy score is : 0.514866199252579 8

MSE = 318566.7325880989

RMSE= 564.4171618475991



MODEL TRAINING

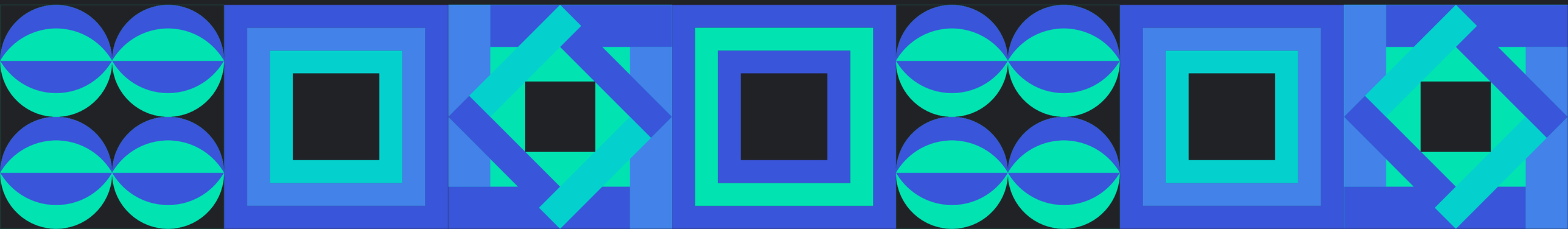
RANDOM FOREST

Model train accuracy score is : 0.9961923371240455

Model test accuracy score is : 0.9798391126381354

MSE = 318566.7325880989

RMSE= 564.4171618475991



conculusion

After See the rsme and accuracy we found the Random Forest best Model

THANK YOU FOR LISTENING!