

Aqar Linear regression model

Abeer AlRuwayti
27/9/2021

Methodology

Web Scraping

EDA

Feature engineering

Linear regression models

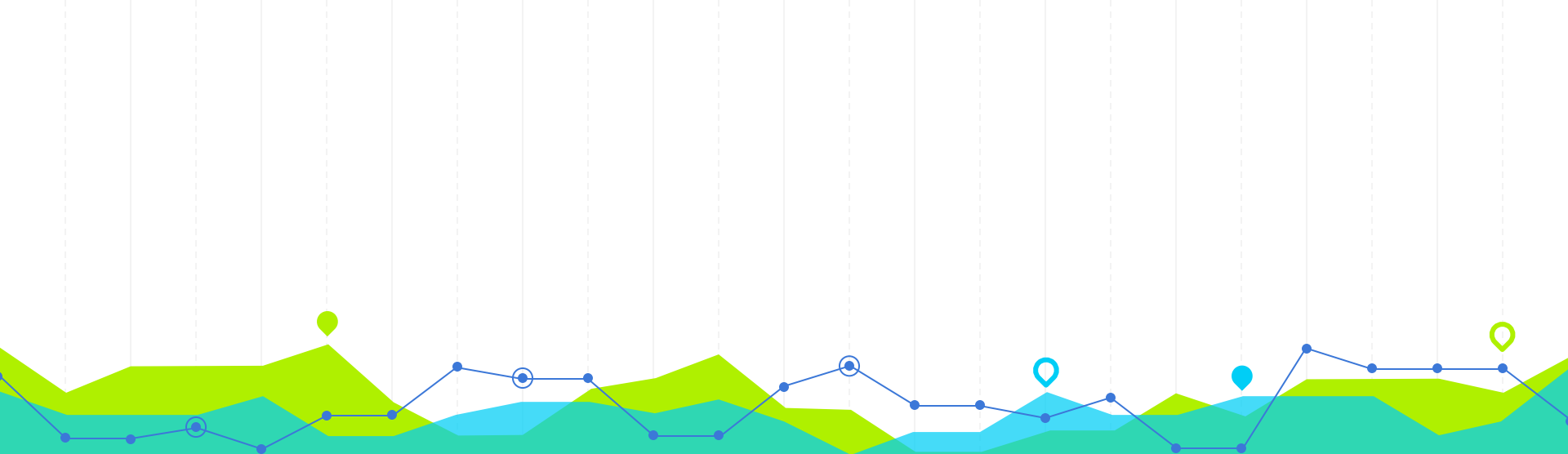
Features

Area, Direction, Bed, Path
Living, Street width, House age, Views

Target

Price





Exploratory data analysis

1

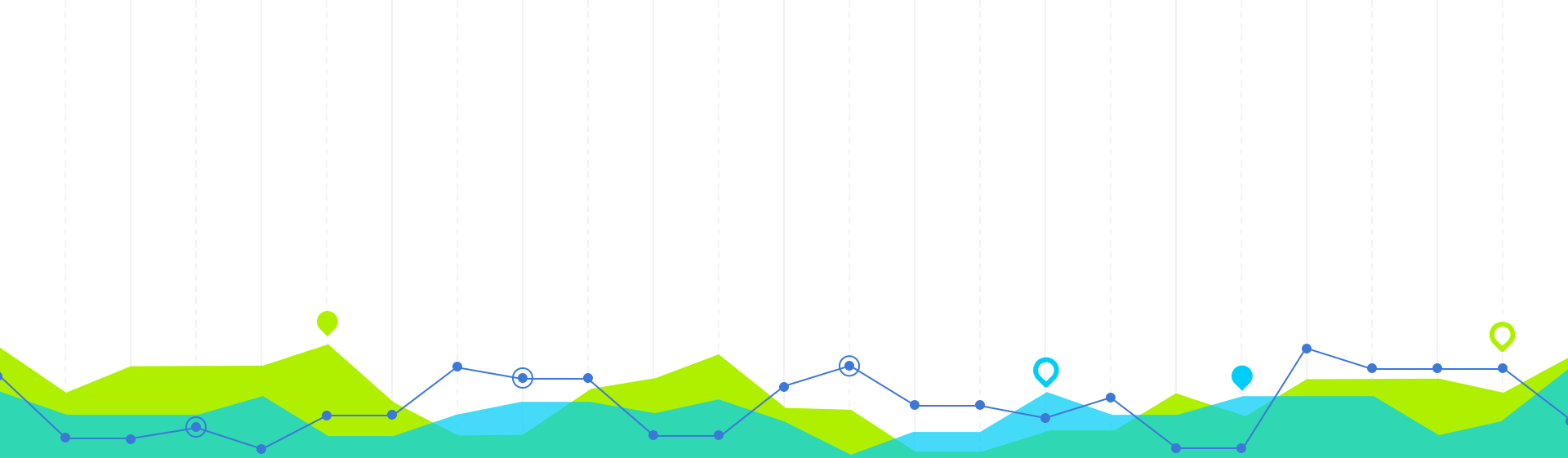
Data cleaning

Missing values (NaNs)

Outliers

Duplicate rows





Feature Engineering

2

Feature engineering



Scaling

The ways to set the feature value range within a similar scale



Dummy variable

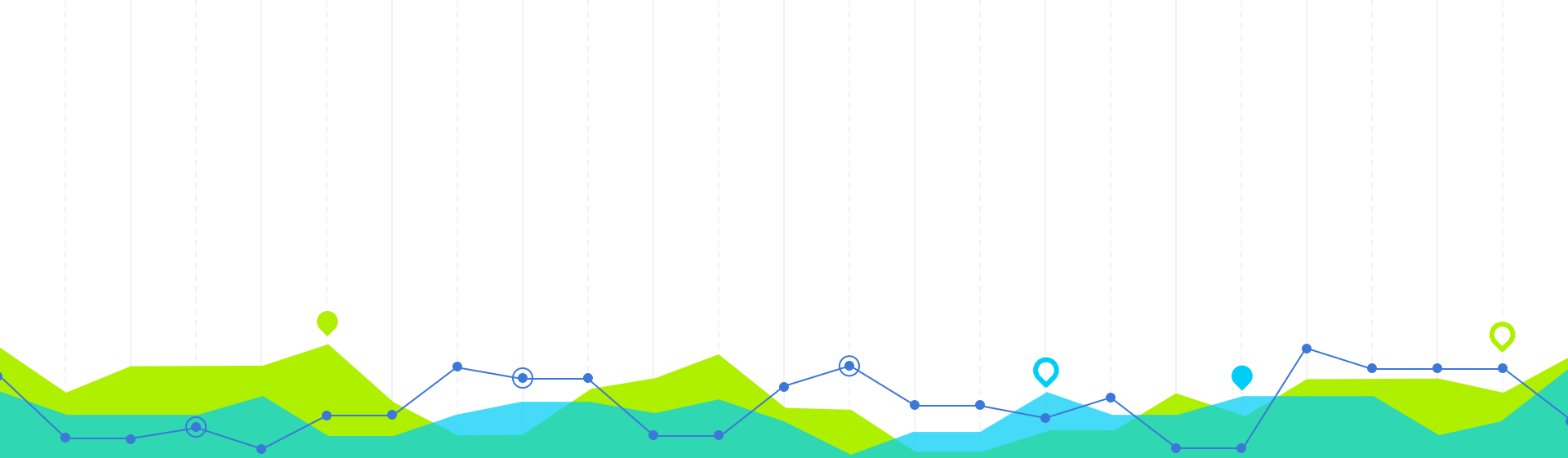
Is a numeric variable that represents categorical data such as direction and house age



Log Transform

Helps to handle skewed data and after transformation, the distribution becomes more approximate to normal.





Linear regression models

3

Models scores

Model	Training	Validation
Baseline model	0.04	0.03
Linear regression	0.1	0.08
Ridge (L1)	0.05	0.06
Lasso model (L2)	0.1	0.09
Polynomial Features	0.1	0.08

THANKS!

Any questions?

