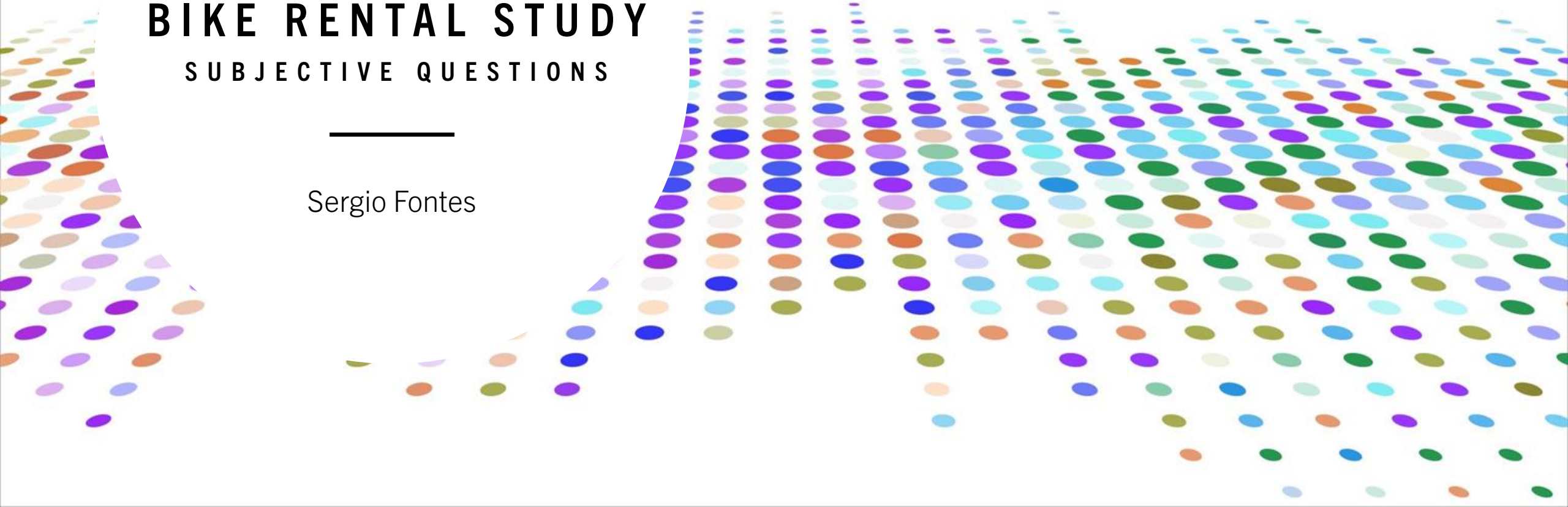
BIKE RENTAL STUDY

SUBJECTIVE QUESTIONS

Sergio Fontes



ASSIGNMENT-BASED SUBJECTIVE QUESTIONS

From your analysis of the categorical variables from the dataset, what could you infer about their effect on the dependent variable?

- They have some effect on the dependent variable.

ASSIGNMENT-BASED SUBJECTIVE QUESTIONS

Why is it important to use drop_first=True during dummy variable creation?

- We want to ensure we don’t have redundant variables on when creating the model

ASSIGNMENT-BASED SUBJECTIVE QUESTIONS

Looking at the pair-plot among the numerical variables, which one has the highest correlation with the target variable?

- “temp”

ASSIGNMENT-BASED SUBJECTIVE QUESTIONS

How did you validate the assumptions of Linear Regression after building the model on the training set?

- dd

ASSIGNMENT-BASED SUBJECTIVE QUESTIONS

Based on the final model, which are the top 3 features contributing significantly towards explaining the demand of the shared bikes?

- “temp”
- “weathersit”
- “season”

GENERAL SUBJECTIVE QUESTIONS

Explain the linear regression algorithm in detail.

- The linear regression algorithm consists in finding the best linear equation that minimizes the error when comparing it to observed values.
- It results on the following equation $y = ax + b + \varepsilon$ where:

  y is the dependent variable

  x is the independent variable or predictor

  b is the constant and a is the ratio that explains how will y increate or decrease per each unit of x

  ε is the associated error of the model.

GENERAL SUBJECTIVE QUESTIONS

Explain the Anscombe's quartet in detail.

- It's a set of four datasets that have identical descriptive statistics but very different distributions.
- They follow the same linear regression model $y = 0{,}5x + 3$
- R2 is the same : 0,67

GENERAL SUBJECTIVE QUESTIONS

What is Pearson's R?

- The Pearson Correlation coefficient is a measure for the linear correlation between 2 variables or sets of data.
- It's a normalized measure and its values lie between -1 and 1.
- If the values are -1 or 1 we have a perfected correlation.
- If the value is 0 no correlation exists.

GENERAL SUBJECTIVE QUESTIONS

What is scaling? Why is scaling performed? What is the difference between normalized scaling and standardized scaling

- Scaling allows us compare variables ensuring variables values don't differ in size of its values.
- With normalized scaling we ensure all values are between [0,1]
- With standardized scaling we also ensure al values are inside the same interval but we ensure the "new" variables have a mean = 0 and standard deviation = 1

GENERAL SUBJECTIVE QUESTIONS

You might have observed that sometimes the value of VIF is infinite. Why does this happen?

If VIF = infinite it means R2 = 1, we have a perfect correlation.

GENERAL SUBJECTIVE QUESTIONS

What is a Q-Q plot? Explain the use and importance of a Q-Q plot in linear regression.

- It's a graphical method to compare 2 distribution based on their quantiles.
- Its always a non-decreasing from left to right
- If both distributions would be identical de Q-Q plot would follow a 45 degree line.
- If the Q-Q plot is approximated to a line we can infer that the data points are normally distributed.