Assignment 3 - Linear regression & forecasting

<u>INSTRUCTIONAL DETAILS</u>

Linear regression is a statistical method used to model the linear relationship between a dependent variable and one or more independent variables. In Python, you can use the LinearRegression class from the sklearn.linear_model module to fit a linear regression model and make predictions.

Here's an example of how you can use the LinearRegression class to fit a linear regression model in Python:

```
from sklearn.linear_model import LinearRegression

# Assume that we have a dataset with two independent variables and one dependent variable
X = [[0, 1], [5, 1], [15, 2], [25, 5], [35, 11], [45, 15], [55, 34], [60, 35]]
y = [4, 5, 20, 14, 32, 22, 38, 43]

# Create a LinearRegression model
model = LinearRegression()

# Fit the model to the data
model.fit(X, y)

# Make predictions using the model
y_pred = model.predict(X)
```

The fit method is used to fit the model to the data, and the predict method is used to make predictions using the fitted model.

You can also use the score method to evaluate the model's performance. For example, to calculate the R-squared value, you can do:

```
r_squared = model.score(X, y)
```

This will give you the coefficient of determination (R-squared) for the model, which is a measure of how well the model fits the data.

The output of the linear regression model is the coefficients and the intercept, which can be used to construct the equation of the line of best fit: y = intercept + coef[0] * x[0] + coef[1] * x[1] + ... + coef[n] * x[n].

You can also use the score() method to evaluate the model's performance on the training data. The returned value is the coefficient of determination (R^2), which is a measure of the model's accuracy. A value of 1 indicates a perfect fit, while a value of 0 indicates that the model is no better than a horizontal line at the mean of the dependent variable.

<u>BACKGROUND</u>

Linear regression is a statistical technique used to analyze the relationship between a dependent variable and one or more independent variables. It is a useful tool for businesses because it allows them to understand how a particular outcome or response is related to a set of predictor variables.

For example, a business might use linear regression to understand how factors such as marketing spend, product quality, and customer satisfaction contribute to overall sales. This can help the business identify which factors have the greatest impact on sales and inform decisions about how to allocate resources in order to maximize sales.

Linear regression can also be used to predict future outcomes based on the relationships identified through the analysis. For example, a business might use linear regression to predict future sales based on expected changes in marketing spend, product quality, and customer satisfaction.

Overall, linear regression is an important tool for businesses because it helps them understand the relationships between different variables and how they contribute to a particular outcome. This can inform decision-making and help businesses achieve their goals.

RESEARCH QUESTION

If Company X built a linear regression model to predict for installations using visitors - how many installations might Company X expect to generate when it acquires 230,000 visitors in a single day?

If Company X built a linear regression model to predict for installations using downloads - how many installations might Company X expect to generate when it acquires 195,000 downloads in a single day?

Data analysis can be a valuable tool in setting corporate key performance indicators (KPIs). A KPI is a metric that is used to measure the performance of an organization or a specific aspect of its operations. Data analysis can help organizations identify the most relevant and meaningful KPIs to track, and can be used to analyze and interpret the data collected on those KPIs.

Here are a few steps that organizations can follow to use data analysis in setting corporate KPIs:

Identify the goals and objectives of the organization: The first step in setting corporate KPIs is to identify the goals and objectives of the organization. This will help determine which KPIs are most relevant and meaningful to track.

Collect and analyze data: The next step is to collect and analyze data on the chosen KPIs. This may involve gathering data from a variety of sources, such as financial reports, customer feedback, and operational data.

Identify trends and patterns: Once you have collected and analyzed the data, you can use data analysis techniques such as visualization and statistical analysis to identify trends and patterns in the data.

Set targets and benchmarks: Based on the trends and patterns identified in the data, you can set targets and benchmarks for the KPIs. These targets and benchmarks can help ensure that the organization is on track to achieve its goals and objectives.

Monitor and adjust: Finally, it is important to regularly monitor the KPIs and make adjustments as needed to ensure that the organization is making progress towards its goals. Data analysis can be used to track the performance of the KPIs and identify areas for improvement.

REQUIREMENTS FOR SUBMISSION
See "Assignment 1 - Descriptives" for a detailed list of submission requirements.

FORMATTING
See "Assignment 1 - Descriptives" for a detailed list of assignment formatting guidelines. Also, assignment formatting guidelines can be found in the course document cache.

DATASET DETAILS (all data sets can be found here or here)

kpisetting.csv

DATASET FIELDS
date
visitors
downloads
installations
28dactive