

Year 10 Digital Technologies Lesson: Python Data Analysis and Visualization

Lesson Overview

This lesson introduces Year 10 students to data analysis and visualization using Python. Students will learn to use pandas for data manipulation, matplotlib for visualization, and basic statistical methods to identify trends and outliers.

Curriculum Alignment

Australian Curriculum: Digital Technologies Content Descriptor: AC9TDI10P02 Analyse and visualise data interactively using a range of software, including spreadsheets and databases, to draw conclusions and make predictions by identifying trends and outliers

Learning Objectives

By the end of this lesson, students will be able to:

1. Import and manipulate data using pandas
2. Create various types of plots using matplotlib
3. Perform basic statistical analysis to identify trends and outliers
4. Draw conclusions and make predictions based on data analysis

Duration

70 minutes

Materials

- Computers with Python, pandas, and matplotlib installed
- Jupyter Notebook or any Python IDE
- Sample dataset (e.g., 'temperature_data.csv' with columns: Date, Temperature)

Lesson Plan

1. Introduction (5 minutes)

- Briefly explain the importance of data analysis in Python
- Introduce the sample dataset and its context

2. Setting Up and Data Import (10 minutes)

- Guide students through importing necessary libraries and the dataset

```
import pandas as pd
import matplotlib.pyplot as plt
```

```
import numpy as np

# Load the data
df = pd.read_csv('temperature_data.csv')
df['Date'] = pd.to_datetime(df['Date'])
print(df.head())
```

3. Basic Data Analysis (15 minutes)

- Demonstrate basic pandas operations for data analysis

```
# Basic statistical analysis
print(df['Temperature'].describe())

# Identifying trends: Moving average
df['MA_7'] = df['Temperature'].rolling(window=7).mean()

# Identifying outliers: Z-score method
z_scores = (df['Temperature'] - df['Temperature'].mean()) /
df['Temperature'].std()
outliers = df[abs(z_scores) > 3]
print("Potential outliers:")
print(outliers)
```

4. Data Visualization (20 minutes)

- Teach students how to create different types of plots using matplotlib

```
# Line plot with moving average
plt.figure(figsize=(12,6))
plt.plot(df['Date'], df['Temperature'], label='Daily Temperature')
plt.plot(df['Date'], df['MA_7'], label='7-day Moving Average', color='red')
plt.scatter(outliers['Date'], outliers['Temperature'], color='green',
label='Outliers')
plt.title('Temperature Trends and Outliers')
plt.xlabel('Date')
plt.ylabel('Temperature (°C)')
plt.legend()
plt.show()

# Histogram for temperature distribution
plt.figure(figsize=(10,5))
plt.hist(df['Temperature'], bins=20, edgecolor='black')
plt.title('Temperature Distribution')
plt.xlabel('Temperature (°C)')
plt.ylabel('Frequency')
plt.show()
```

5. Drawing Conclusions and Making Predictions (15 minutes)

- Guide students in interpreting their visualizations and analysis results
- Introduce simple linear regression for prediction

```
# Simple linear regression for prediction
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split

X = df.index.values.reshape(-1, 1)
y = df['Temperature'].values

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)

model = LinearRegression()
model.fit(X_train, y_train)

# Predict next 7 days
future_dates = np.array(range(len(df), len(df)+7)).reshape(-1, 1)
future_temps = model.predict(future_dates)

print("Predicted temperatures for the next 7 days:")
print(future_temps)
```

6. Reflection and Discussion (5 minutes)

- Discuss the advantages and limitations of the analysis methods used
- Briefly introduce more advanced data analysis techniques in Python

Hands-on Activity

Throughout the lesson, students will work with the sample dataset to:

1. Perform basic statistical analysis using pandas
2. Create line plots and histograms to visualize temperature trends
3. Identify outliers using the Z-score method
4. Use simple linear regression to predict future temperatures

Worksheet Template

1. What is the average temperature in the dataset? How does this compare to the median temperature?
2. Create a line plot of the temperatures over time. What trends do you observe?
3. Using the Z-score method, how many outliers did you identify? What might these outliers represent?
4. Based on the histogram, how would you describe the distribution of temperatures?

5. Using the linear regression model, what is the predicted temperature for 7 days from the last date in our dataset?
6. What conclusions can you draw about temperature trends from your analysis? How reliable do you think your predictions are?

Assessment

- Completed worksheet
- Quality of data visualizations created
- Ability to interpret results and draw conclusions
- Participation in class discussion

Extension Activities

- Analyze a more complex dataset with multiple variables
- Implement more advanced outlier detection methods
- Create an interactive dashboard using libraries like Plotly