CoGrammar

# DS PORTFOLIO SESSION 10

SKILLS FOR LIFE
SKILLS BOOTCAMPS

Department for Education

# Data Science Session Housekeeping

- The use of disrespectful language is prohibited in the questions. This is a supportive, learning environment for all - please engage accordingly. **(FBV: Mutual Respect.)**
- No question is daft or silly - **ask them!**
- There are **Q&A sessions** midway and at the end of the session, should you wish to ask any follow-up questions. Moderators are going to be answering questions as the session progresses as well.
- If you have any questions outside of this lecture, or that are not answered during this lecture, please do submit these for upcoming Open Classes. You can submit these questions here: **Open Class Questions**

# **Data Science Session Housekeeping** cont.

- For all **non-academic questions**, please submit a query: **www.hyperiondev.com/support**
- Report a **safeguarding** incident: **www.hyperiondev.com/safeguardreporting**
- We would love your **feedback** on lectures: **Feedback on Lectures**

# Progression Criteria

### ✅ Criterion 1: Initial Requirements

- Complete 15 hours of Guided Learning Hours and the first four tasks within two weeks.

### ✅ Criterion 2: Mid-Course Progress

- Software Engineering: Finish 14 tasks by week 8.
- Data Science: Finish 13 tasks by week 8.

### ✅ Criterion 3: Post-Course Progress

- Complete all mandatory tasks by 24th March 2024.
- Record an Invitation to Interview within 4 weeks of course completion, or by 30th March 2024.
- Achieve 112 GLH by 24th March 2024.

### ✅ Criterion 4: Employability

- Record a Final Job Outcome within 12 weeks of graduation, or by 23rd September 2024.

# Recap of Week 10: Data Visualization & Analysis

## Approach to Visualisation

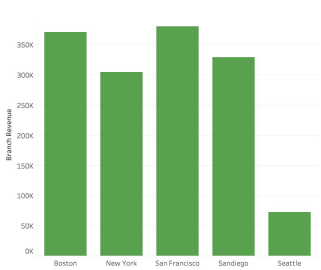
1. Start with a processed and clean dataset.
2. Know your dataset.
3. Determine what you want to find.
4. Create data visualisations.
5. Refine your visualisation.
6. Note down your findings.

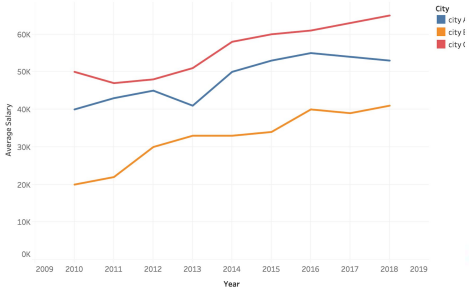
## Types of Data:

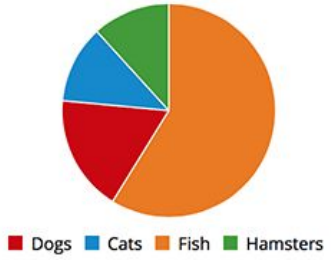
- Discrete
- Categorical
- Continuous
- Time Series

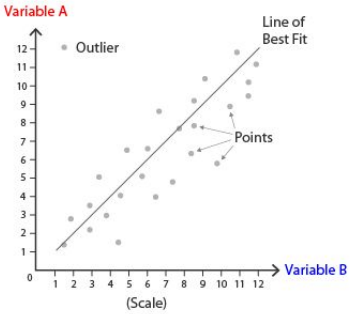
## Types of Data:

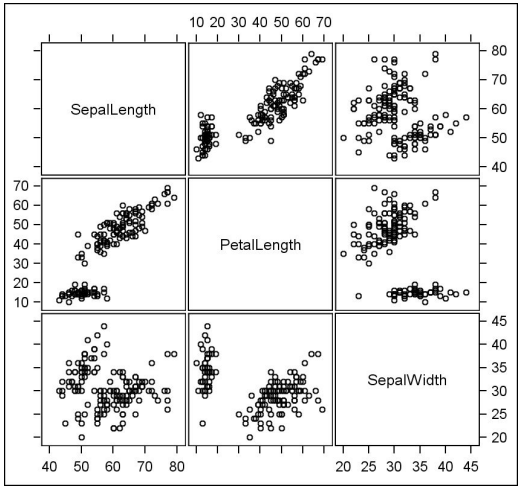
- Bar Chart
- Line Graphs
- Pie Chart
- Scatterplot
- Scatterplot Matrix
- Double Axis Chart




# FutureTech Industries

- **Background:** Enter "FutureTech Industries", a leading manufacturing company in Technopolis. They've recently faced a series of unexpected machinery breakdowns, causing production halts and financial setbacks. The CEO of FutureTech, Dr. Ada Lovelace, believes that there's a solution in the realm of machine learning.

- **Challenge:** Temperature readings, vibration levels, operational hours, and more. Your mission? To unravel the mysteries within this data and develop a machine learning model that can predict when a machine is on the brink of failure. By doing so, you can help FutureTech transition from **reactive** repairs to **proactive** maintenance, ensuring the gears of Technopolis never stop turning.

- **Objective:** Develop a machine learning model that can predict when a machine is on the brink of failure.

# What is the main challenge that FutureTech Industries is facing, and how do you think Dr. Ada Lovelace believes machine learning can help?

A. Managing human resources efficiently
B. Dealing with unexpected machinery breakdowns by predicting failures using machine learning
C. Implementing advanced robotics for manufacturing
D. Enhancing the aesthetic appeal of machinery

# Demo: Predictive maintenance

```
# Simple example to demonstrate the need for predictive maintenance in
FutureTech Industries

import pandas as pd

from sklearn.model_selection import train_test_split

from sklearn.ensemble import RandomForestClassifier

from sklearn.metrics import accuracy_score


# Assume machinery_data is a Pandas DataFrame with sensor data

machinery_data = pd.DataFrame({

  'Temperature': [45, 50, 55, 60, 65, 70],

  'Vibration_Level': [0.2, 0.5, 0.8, 1.2, 1.5, 1.8],

  'Operational_Hours': [500, 600, 700, 800, 900, 1000],

  'Failure_Status': [0, 0, 0, 1, 1, 1] # 0 represents normal, 1 represents
failure

})
```

```
# Split the data into features (X) and target variable (y)

X = machinery_data[['Temperature', 'Vibration_Level', 'Operational_Hours']]

y = machinery_data['Failure_Status']


# Split the data into training and testing sets

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


# Create a random forest classifier

clf = RandomForestClassifier()


# Train the classifier

clf.fit(X_train, y_train)
```

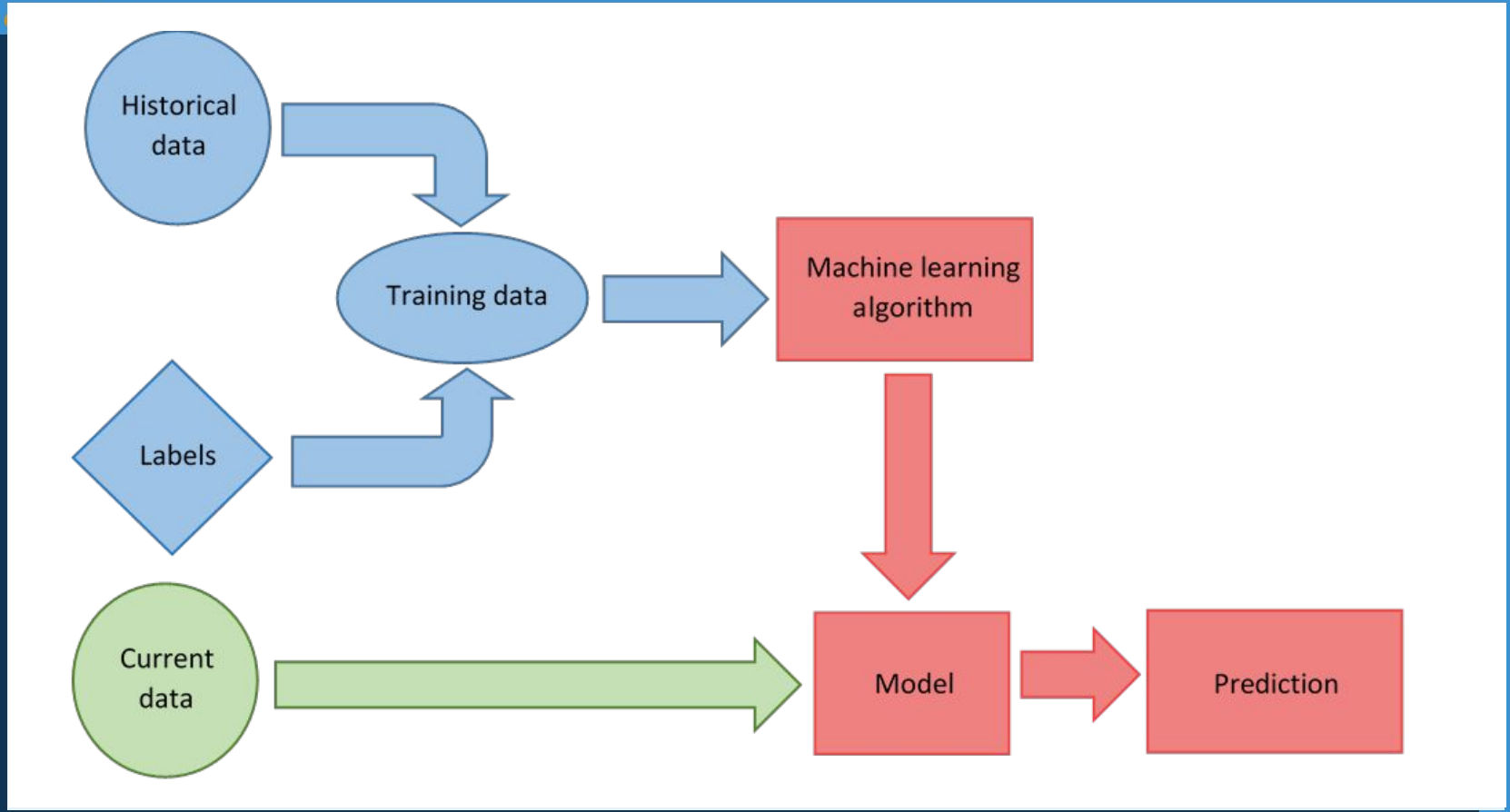
```
# Make predictions on the test set

y_pred = clf.predict(X_test)


# Evaluate the accuracy of the model

accuracy = accuracy_score(y_test, y_pred)


print(f"Accuracy of the predictive maintenance model: {accuracy}")
```




https://www.plass.com/sites/default/files/inline-images/des.png

# Demo: Advanced Machine Learning

```
# Extended example to demonstrate advanced machine learning for predictive
maintenance
# (Demonstration may include additional features not covered in the example)

from sklearn.metrics import classification_report, confusion_matrix

# Assume additional sensor data is available for training the model
additional_machinery_data = pd.DataFrame({
    'Temperature': [75, 80, 85, 90, 95],
    'Vibration_Level': [2.0, 2.2, 2.5, 2.8, 3.0],
    'Operational_Hours': [1100, 1200, 1300, 1400, 1500],
    'Failure_Status': [1, 1, 1, 0, 0]
})

# Concatenate the additional data with the original data
machinery_data = pd.concat([machinery_data, additional_machinery_data],
ignore_index=True)

# Split the combined data into features (X) and target variable (y)
X_combined = machinery_data[['Temperature', 'Vibration_Level',
'Operational_Hours']]
y_combined = machinery_data['Failure_Status']
```

```
# Create a new random forest classifier
clf_combined = RandomForestClassifier()

# Train the new classifier on the combined data
clf_combined.fit(X_combined, y_combined)

# Make predictions on the original test set
y_pred_combined = clf_combined.predict(X_test)

# Evaluate the accuracy of the combined model
accuracy_combined = accuracy_score(y_test, y_pred_combined)

# Display additional evaluation metrics
classification_rep = classification_report(y_test, y_pred_combined)
confusion_mat = confusion_matrix(y_test, y_pred_combined)

print(f"Accuracy of the combined predictive maintenance model:
{accuracy_combined}")
print("\nClassification Report:")
print(classification_rep)
print("\nConfusion Matrix:")
print(confusion_mat)
```

# Demo: Predictive Maintenance (Output)

```
Accuracy of the combined predictive maintenance model: 1.0

Classification Report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00         2

    accuracy                           1.00         2
   macro avg       1.00      1.00      1.00         2
weighted avg       1.00      1.00      1.00         2


Confusion Matrix:
[[2]]
```

# FutureTech

Enter "FutureTech Industries", a leading manufacturing company in Technopolis. They've recently faced a series of unexpected machinery breakdowns, causing production halts and financial setbacks

Helpful sklearn modules/classes:

| |
|---|
| RandomForestClassifier |
| AccuracyScore |
| ClassificationReport |
| ConfusionMatrix |

## Important Concepts:

1. **Machine Learning Forms:** Supervised learning can be used to implement the functionality that determines the ‘failure status’ of the machines by training the models using variables such as ‘temp’, ‘vibration level’ and ‘operational hours’.
2. **Random Forest:** This classifier can be trained on the training set to make predictions on the test set.
3. **Evaluation Metrics:** Make use of classification reports and confusion matrices to assess the performance of the predictive maintenance model.

## Advanced Challenge:

- Assess and address imbalances found in the dataset between ‘normal’ and ‘failure’ states.
- Experiment by oversampling, undersampling, or using different class weights, to handle imbalances

# In the machine learning model for predictive maintenance, what does the 'Failure_Status' variable represent?

A. The temperature of machinery.
B. The vibration level of machinery.
C. Operational hours of machinery.
D. Whether the machinery is in a normal state or on the brink of failure.

# Summary

## Machine Learning

★ A branch of AI that focuses on creating systems that learn from data, enabling software to improve its performance over time.

## Supervised Learning with Random Forest

★ Supervised learning is where the model is trained on labeled data
★ The Random Forest algorithm is an ensemble-learning method that can be used for predictive maintenance.

# Questions and Answers

## Questions around the Case Study