Answers

1. **A) Create two series as shown using pd.series() function.Series_A = [10,20,30,40,50] Series_B = [40,50,60,70,80]. Get the items common to both.**

    Program:
    ```
    import pandas as pd
    import numpy as np

    x = pd.Series([10, 20, 30, 40, 50])
    y = pd.Series([40, 50, 60, 70,80])

    intersect = pd.Series(np.intersect1d(x, y))

    print("Items that are common.")
    print(intersect)
    ```

    Output:
    ```
    Items that are common.
    0    40
    1    50
    dtype: int64
    ```

    B) Create a data frame with following data

Ename	Type	Dname	exp	salary
Roshan	regular	cs	10	50000
Amar	adhoc	cs	20	15000
Ashwini	regular	ec	5	30000
Lohith	adhoc	ec	14	15000
Mohan	contract	cs	9	10000
Pramod	regular	ec	8	40000

    i) **Make pivot table that shows the average salary of each employee for each department.**

        Program:
        ```
        import pandas as pd

        df=pd.DataFrame({"Ename": ["Roshan","Amar","Ashwini","Lohith","Mohan","Pramod"],
                "Type":["Regular","adhoc","Regular","Contract","Regular","Regular"],
                "Dname":["cs","cs","ec","ec","cs","ec"],
                "exp":[10,20,5,14,9,8],
                "Salary":[50000,15000,30000,15000,10000,40000]})

        print("\nDataset Created\n")
        print(df)
        print("\n")

        pivot=df.pivot_table(index=['Dname'],values=['Ename','Salary'],aggfunc=['mean'])
        print(pivot)
        ```

Answers

Output:

```
Dataset Created

Ename      Type Dname  exp  Salary
0   Roshan   Regular   cs   10   50000
1     Amar     adhoc   cs   20   15000
2  Ashwini   Regular   ec    5   30000
3   Lohith  Contract   ec   14   15000
4    Mohan   Regular   cs    9   10000
5   Pramod   Regular   ec    8   40000


             mean
           Salary
Dname
cs      25000.000000
ec      28333.333333
```

ii) **Make a pivot table that shows the sum and mean of the salaries of each type of employee and the number of employees of each type.**
Program:
pivot =df.pivot_table(index=['Type'],values=['Salary'],aggfunc={'mean','sum',})
print(pivot)

Output:
```
            mean      sum
          Salary   Salary
Type
Contract  15000    15000
Regular   32500   130000
adhoc     15000    15000
```

iii) **You have come across some missing data in your data set. How will you handle it?**

When dealing with missing data, data scientists can use 2 primary methods to solve the error.

1. Imputation
2. Removal of Data

1. Imputation

The Imputation method develops reasonable guesses for missing data. It is most useful when the percentage of missing data is low. Depending why the data are missing, imputation methods can deliver reasonably reliable results.

Mean, Mode and Median
This is one of the most common methods of imputing values when dealing with missing data. In cases, where there are small number of missing observations, data scientist can calculate mean or median of the existing observation.

Answers

2. Removal of Data

   When dealing with data that is missing at random, related data can be deleted to reduce bias. There are 2 primary methods while dealing with deletion of data. They are

   1. Listwise

      In this method, all data for an observation that has one or more missing values are deleted. The analysis, is run only on observations that have a complete set of data. If the data set is small, it may be the most efficient method to eliminate those cases from the analysis. However, in most cases this is used when, the data are not missing completely at random.

   2. Pairwise

      Pairwise deletion saves more data compared to likewise deletion because the former only deletes entries where variables were necessary for testing

iv) To start Linear Regression, you would need to make some assumptions. What are those?

The regression has five key assumptions:
- Linear relationship.
- Multivariate normality.
- No or little multicollinearity. (Multicollinearity is a statistical concept where several independent variables in a model are correlated.)
- No auto-correlation.
- Homoscedasticity.

It consists of 3 stages – (1) analyzing the correlation and directionality of the data, (2) estimating the model, i.e., fitting the line, and (3) evaluating the validity and usefulness of the model.
The two things should be done before one performs a regression analyst. Well, the first thing you'll need to do is one construct a scatter plot because that imply the correlation coefficient. And the second thing you need to do is that you need to test the significance of the relationship between the two rows of data.

v) How do you clean up & organize large datasets? Explain briefly?

Step 1: Remove duplicate or irrelevant observations.
      Remove unwanted observations from your dataset, including duplicate observations or irrelevant observations.
Step 2: Fix structural errors. ...
Step 3: Filter unwanted outliers. ...
Step 4: Handle missing data. ...
Step 5: Validate and QA.

Answers

vi) How does Netflix use Machine learning?

There are many applications of Machine Learning used by Netflix among which the most popular is their content recommendation system that tells you what you might like. Netflix uses machine learning in almost all the facets of its working to provide a seamless experience for user s All this data can be used for machine learning that ultimately improves the bottom line i.e. gets more subscribers for Netflix! So let's check out the different ways Netflix uses Machine Learning

1. **Content Recommendations**

   They use their recommendations system that is based on a machine-learning algorithm that takes into account your past choices in movies, the types of genres you like, and what moves were watched by users that had similar tastes like yours. This movie recommendation algorithm is very important for Netflix, as they have thousands of options of all types and users, are more likely to get confused in choosing what to watch next than actually watching anything

2. **Auto-generated Thumbnails**

   Netflix uses machine learning to analyze your movie and series choices and understand what sort of thumbnail you are most likely to click. For example, the series Riverdale can have two different thumbnails, a sweet romantic version or a serious mystery version and you will see the one depending on your tastes.

3. **Streaming Quality**

   Netflix uses machine learning algorithms to predict the viewer patterns and understand when there will be general increases and decreases in viewers of spikes in viewing a certain movie or show. Then they can cache the regional servers that are much closer to the viewers so that there is no log in streaming or loading times even during peak popularity periods

4. **Locations for Shooting**

   All these factors need to be considered by the production team before deciding on a location and machine learning can be a big help in this. A machine learning algorithm can be used to find a list of optimal locations in the world based on all the constraints that need to be satisfied.

5. **Content Quality**

   they use Machine Learning for this as well! Netflix has created a supervised quality control algorithm that passes or fails the content such as audio, video, subtitle text, etc. based on the data it was trained on. If any content is failed, then it is further checked by manually quality control to ensure that only the best quality reached the users. After all, you probably wouldn't watch Stranger Things on Netflix if the subtitles were wrong or the audio was lagging behind the video

Answers

vii) Explain workflow in machine learning using a day-to-day life example with diagram?
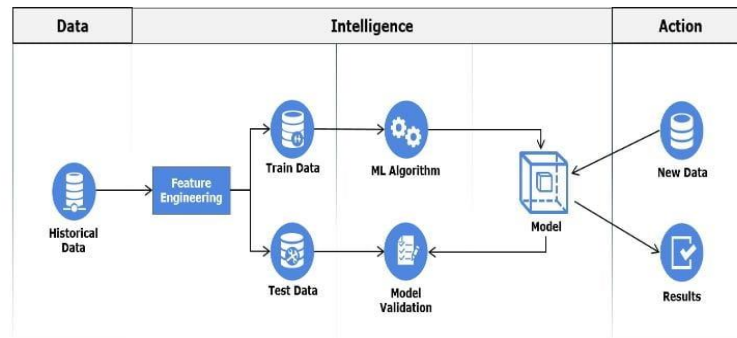
In this process, first relevant data is gathered then it is cleaned and transformed through a process called Feature Engineering. During the process of Feature Engineering, handling missing value, handling outliers, creating new features out of new existing ones are some of the common tasks performed.

After feature engineering, the data is split into train data and Test data. The train data is used for training the machine learning model. Once the model is built, it is validated against the Test data for accuracy. This accuracy help up in estimating the performance on previously unseen data. If the model performance on both Train and Test Data is satisfactory, the model may be deployed.

Once deployed, the model make predictions- on new data; this predictions are used to take business decisions.

Some of the examples are:

1.Virtual Personal Assistants
Virtual Assistants are integrated to a variety of platforms. For example:
- Smart Speakers: Amazon Echo and Google Home
- Smartphones: Samsung Bixby on Samsung S8
- Mobile Apps: Google Allo

2. Predictions while Commuting Traffic Predictions:
We all have been using GPS navigation services. While we do that, our current locations and velocities are being saved at a central server for managing traffic. This data is then used to build a map of the current traffic.

3. Videos Surveillance :
the video surveillance system nowadays is powered by AI that makes it possible to detect crime before they happen. They track unusual behaviour of people like standing motionless for a long time, stumbling, or napping on benches etc

4. Home Security and Smart Homes
AI-powered alarms and cameras are now at the forefront of cutting edge home security. These security systems use facial recognition software and machine learning models to build a catalogue of your home's frequent visitors. This allows the system to detect uninvited guests.

Answers

2   A) Describe the different types of machine learning algorithms with examples.

Types of Machine Learning algorithms
- o   Supervised Machine Learning
- o   Unsupervised Machine Learning
- o   Reinforcement Learning

1.Supervised Machine Learning.

The main goal of the supervised learning technique is to map the input variable(x) with the output variable(y). Some real-world applications of supervised learning are Risk Assessment, Fraud Detection, Spam filtering, etc.

There are 2 categories of algorithm.

a.Classification
The classification algorithms predict the categories present in the dataset. Some real-world examples of classification algorithms are Spam Detection, Email filtering, etc.
b. regression
Regression algorithms are used to solve regression problems in which there is a linear relationship between input and output variables. These are used to predict continuous output variables, such as market trends, weather prediction, etc.

2. Unsupervised Learning
The main aim of the unsupervised learning algorithm is to group or categories the unsorted dataset according to the similarities, patterns, and differences.

There are 2 categories in unsupervised learning algorithms
Association
Association rule learning is an unsupervised learning technique, which finds interesting relations among variables within a large dataset.
Clustering
The clustering technique is used when we want to find the inherent groups from the data.

3. Reinforcement learning
   Reinforcement learning works on a feedback-based process, in which an AI agent (A   software component) automatically explore its surrounding by hitting & trail, taking action, learning from experiences, and improving its performance.

B) What are the goals of artificial intelligence?

1. Develop problem-solving ability.
2. Incorporate knowledge representation.
3. Facilitate planning
4. Allow continuous learning
5. Encourage social Intelligence
6.Promote Creativity.
7. Achieve general intelligence
8. Promote synergy between humans and AI

Answers

3A) Use'Cars93' dataset to answer the below questions. The information that the columns of this dataset contain is given below:

Manufacturer	Model	Type	Price	MPG.city	MPG.highway	Horsepower	Rear.seat.room	Passengers
Manufacturer.	Model.	Type: a factor with levels "Small", "Sporty", " Compact", "Midsize", "Large" and "Van".	Midrange Price (in \$1,000).	City MPG (miles per US gallon by EPA rating).	Highway MPG.	Horsepower (maximum).	Rear seat room (inches) (missing for 2- seater vehicles).	Passenger capacity (persons)

Create the following plots to visualize/summarize the data and customize it appropriately.

i)   **Use a box plot to determine the price range of all different cars available in the market? And interpret the five-number summary**

Program:

```
import matplotlib
from matplotlib import pyplot as plt
import seaborn as sns
import pandas as pd
import math
import numpy as np

df=pd.read_csv("cars420.csv")
df=df.head()
fig=plt.figure()
plt.title("Price Range of Car")
sns.boxplot(df["Price"])

#Interpreting the five number Summary
print(np.min(df.Price))
print(np.max(df.Price))
print(np.std(df.Price))
print(np.mean(df.Price))
print(np.median(df.Price))
```
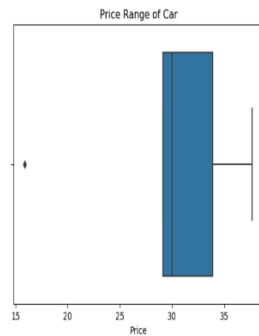
Output:

```
15.9
37.7
7.373031940796134
29.320000000000004
30.0
```

Answers

Price Range of Car

ii) **Histogram to check the frequency distribution of the variable 'Mpg.city' (Miles per gallon) and note down the interval having the highest frequency.**

Program:

```
import matplotlib
from matplotlib import pyplot as plt
import pandas as pd
import numpy as np

df=pd.read_csv("cars420.csv")

mode1=df['MPG_city'].mode()
print("Highest frequency : ",mode1)

fig=plt.figure()
x=df.MPG_city
plt.title("Frequency distribution")
plt.xlabel("MPG_city")
plt.hist(x, bins=5, edgecolor="black")
```
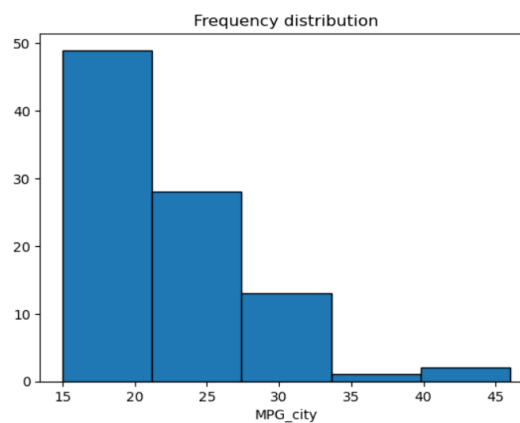
Output:

```
Highest frequency :  0    18
```


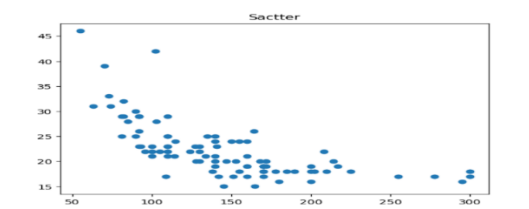Frequency distribution

Answers

- **Use a scatter plot to determine whether a car with higher horsepower gives lower mileage?**

  Program:

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

  fig=plt.figure()
  x=df.Horsepower
  y=df.MPG_city
  plt.title("Sactter")
  plt.scatter(x,y)
  ```

  Output:

- **Use a line chart to observe the variations in 'Engine Size', against 'Horsepower'.**
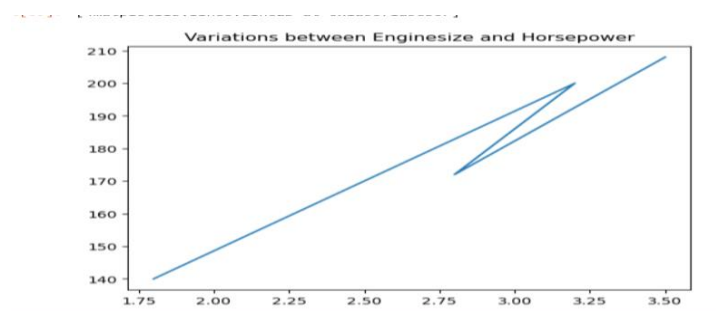
  Program:

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

  df=df.head()

  fig=plt.figure()
  x=df.EngineSize
  y=df.Horsepower
  plt.title("Variations between Enginesize and Horsepower")
  plt.plot(x,y)
  ```
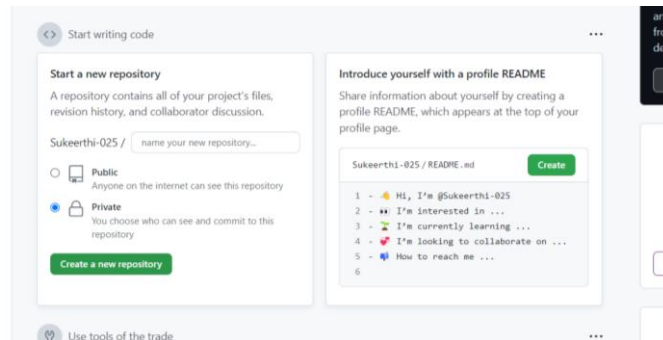
  Output:

Answers

- **Create a git repository and push source code to the repository**

  a) Save the above code in a format. Now, Open the github in browser and login using ur email id and click ok "Create Repository"

  b) Enter the new Repository name.

  c) Click on create new repository.

  d) Click on upload existing file.
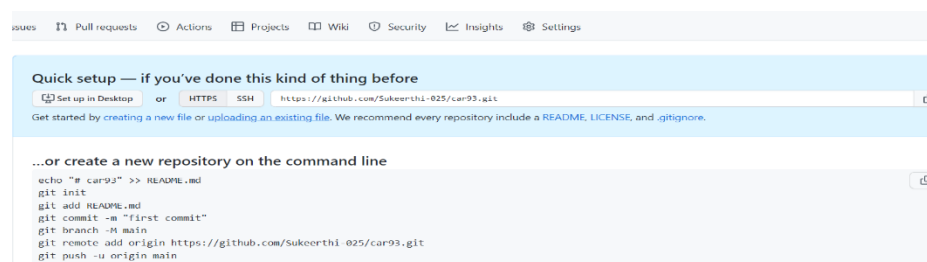
Answers

e) Nw, click on choose your file and select the file you stored your source code and commit changes.

car93 /

Drag additional files here to add them to your repository
Or choose your files

source code.pdf
Uploading 1 of 1 files

Commit changes

Add files via upload

Add an optional extended description...

Commit changes   Cancel

f) Nw you have created the git repository.

Sukeerthi-025 / **car93** Public   Pin   Unwatch 1 ▾   Fork 0 ▾   Star 0

<> Code   Issues   Pull requests   Actions   Projects   Wiki   Security   Insights   Settings

main ▾   1 branch   0 tags   Go to file   Add file ▾   <> Code ▾   **About**

Sukeerthi-025 source code.commit   34891b0 now   1 commit   No description, website, or topics provided.

source code.pdf   source code.commit   now   0 stars
                                                                                   1 watching
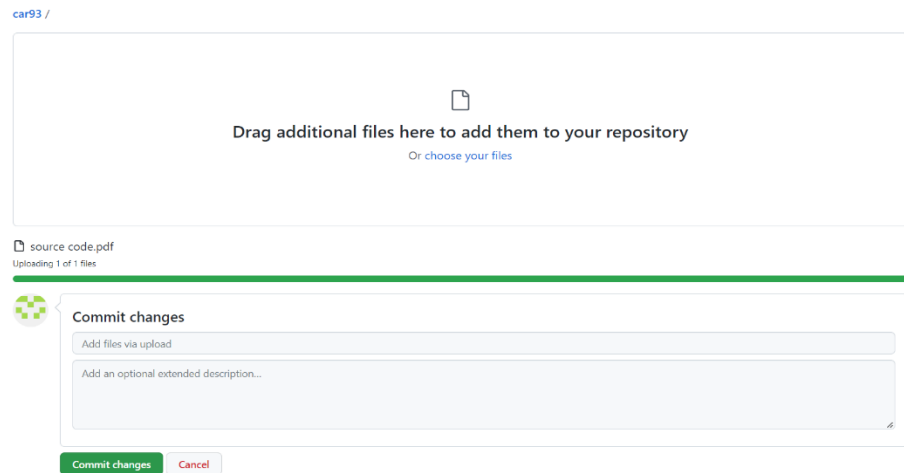Help people interested in this repository understand your project by adding a README.   Add a README   0 forks

                                                                                        Releases

                                                                                        No releases published
                                                                                        Create a new release

                                                                                        Packages

                                                                                        No packages published
                                                                                        Publish your first package

3b.

1. **find a list of squares of the first five odd numbers using lambda and map function.**

   Program:

   ```
   b=(1,3,5,7,9)
   odd=list(map(lambda a : a*a,b))
   print(odd)
   ```
   Output:

   [1, 9, 25, 49, 81]

2. **find the odd numbers from a given list using a filter**

   Program:

   ```
   a=(5,6,7,8,9)
   odd=list(filter(lambda a:a%2!=0,a))
   print(odd)
   ```

Answers

Output:

```
[5, 7, 9]
```

3. compute a sum of the first five integers using reduce function.

```
from functools import reduce
c=[0,1,2,3,4]
print(reduce(lambda a,b:a+b,c)
```

Output:

```
10
```

1. if an organization needs to run highly sensitive & mission critical application with unpredictable performance & capacity Requirements, which model would be best to meet its requirements

Ans:

Hybrid cloud is IT infrastructure that connects at least one public cloud and at least one private cloud, and provides orchestration, management and application portability between them to create a single, flexible, optimal cloud environment for running a company's computing workloads

2. Is it possible to run completely different operating systems on virtual machines that are on single host? If yes, what makes this possible

Ans:

Yes
Virtual box

3. Cloud adoption is an integral part of application modernization. What are the other two important components of modernization?

Ans:

Cloud Adoption is a strategic move by organisations of reducing cost, mitigating risk and achieving scalability of data base capabilities. Cloud adoption may be up to various degrees in an organisation, depending on the depth of adoption. In fact the depth of adoption yields insight into the maturity of best practices, enterprise-ready cloud services availability.

☐ the two important components of modernization are microservices and DevOps microservices- Microservices - also known as the microservice architecture - is an architectural style that structures an application as a collection of services that are

• Highly maintainable and testable
• Loosely coupled
• Independently deployable
• Organized around business capabilities
• Owned by a small team

DevOps - DevOps is the combination of cultural philosophies, practices, and tools that increases an organization's ability to deliver applications and services at high velocity: evolving and improving products at a faster pace than organizations using traditional software development

Answers

and infrastructure management processes. This speed enables organizations to better serve their customers and compete more effectively in the market