GOVERNMENT POLYTECHNIC NAGAMANGALA

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Vth Semester Diploma

**Artificial Intelligence and Machine Learning (20CS51)**

**Assignment: 01**

**NAME: RASHMI**

**ROLL NO: 158CS22046**

AIML (20CS51)

ASSIGNMENT – WEEK 02

1. Download any two datasets from the internet and perform the following operations.

1. Aggregate functions.
2. Use Map, Filter, Reduce, and Lambda Functions with Pandas data frames
3. Visualize the data set (At least 6 different plots).
4. How do you create a project plan and product backlog for an AI project? (Everyone chooses the area you want to work on or do the research work. Give a brief introduction to the work carried out and the final report to be submitted at the end of the course.)
   1. Create a Git Repository for following the Regression Project ML / deep learning.
   2. Classification Project – ML / deep learning
   3. Clustering project – ML / deep learning
   4. Natural Language Processing – ML / deep learning

**Important Note:**

1. **Last Date for Submission:17-07-2024.**
2. **Everyone must perform the above operation using different datasets.**
3. **Submit the report to the email** [**aimlgtn@gmail.com**](mailto:aimlgtn@gmail.com)

**1. Download any two datasets from the internet and perform the following operations.**

**DATASET:01(**/content/student-dataset.csv**)**

**DESCRIPTION**

* id
* name
* nationality
* city
* latitude
* longitude
* gender
* ethnic.group
* age
* english.grade
* math.grade
* sciences.grade
* language.grade
* portfolio.rating
* coverletter.rating
* refletter.rating

**(a) Aggregate functions.**

* HEAD

import pandas as pd

path="/content/student-dataset.csv"

df=pd.read\_csv(path)

df.head(5)

OUTPUT

id name nationality city latitude longitude gender ethnic.group age english.grade math.grade sciences.grade language.grade portfolio.rating coverletter.rating refletter.rating

0 0 Kiana Lor China Suzhou 31.31 120.62 F NaN 22 3.5 3.7 3.1 1.0 4 4.0 4

1 1 Joshua Lonaker United States of America Santa Clarita 34.39 -118.54 M NaN 22 2.9 3.2 3.6 5.0 5 4.0 5

2 2 Dakota Blanco United States of America Oakland 37.80 -122.27 F NaN 22 3.9 3.8 3.2 5.0 3 3.0 4

3 3 Natasha Yarusso United States of America Castro Valley 37.69 -122.09 F NaN 20 3.3 2.8 3.2 5.0 5 2.0 4

4 4 Brooke Cazares Brazil São José dos Campos -23.18 -45.88 F NaN 21 3.7 2.6 3.4 1.0 4 4.0 5

* **TAIL**

df.tail(10)

OUTPUT

|  | **name** | **nationality** | **city** | **latitude** | **longitude** | **gender** |  | **ethnic.group** | **age** | **english.grade** | **math.grade** | **sciences.grade** | **language.grade** | **portfolio.rating** | **coverletter.rating** | **refletter.rating** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **297** | 297 | **id** | United States of America | Charlotte | 35.23 | -80.84 |  | M | NaN | 23 | 3.8 | 3.5 | 3.9 | 5.0 | 5 | 2.0 | 5 |
| **298** | 298 | Nicolas Papp | United States of America | New York | 40.71 | -74.01 |  | M | NaN | 23 | 2.7 | 3.7 | 3.9 | 5.0 | 5 | 4.0 | 4 |
| **299** | 299 | Christopher Inhulsen | United States of America | Fort Worth | 32.73 | -97.32 |  | M | NaN | 23 | 3.1 | 3.3 | 2.9 | 5.0 | 2 | 4.0 | 4 |
| **300** | 300 | Tessa Rawanduzy | United States of America | Derby | 37.55 | -97.27 |  | F | NaN | 21 | 3.3 | 3.3 | 2.0 | 5.0 | 3 | 3.0 | 5 |
| **301** | 301 | Vanessa Gonzalez | Nicaragua | León | 12.44 | -86.88 |  | F | NaN | 24 | 3.3 | 3.7 | 3.8 | 3.0 | 5 | 5.0 | 4 |
| **302** | 302 | Austin Haas | United States of America | Columbus | 39.96 | -83.00 |  | M | NaN | 20 | 3.6 | 3.7 | 3.1 | 5.0 | 4 | 5.0 | 5 |
| **303** | 303 | Madison Fithian | United States of America | Los Angeles | 34.05 | -118.24 |  | F | NaN | 20 | 3.6 | 3.9 | 4.0 | 5.0 | 5 | 5.0 | 3 |
| **304** | 304 | Zachary Mulvahill | United States of America | Los Angeles | 34.05 | -118.24 |  | M | NaN | 20 | 3.2 | 3.4 | 3.9 | 5.0 | 5 | 5.0 | 3 |
| **305** | 305 | Eliana Michelsen | United States of America | Oakland | 37.80 | -122.27 |  | F | NaN | 23 | 3.0 | 2.8 | 2.9 | 5.0 | 4 | 4.0 | 5 |
| **306** | 306 | Dane Whittemore | Canada | Toronto | 43.67 | -79.42 |  | M | NaN | 21 | 3.8 | 3.2 | 4.0 | 5.0 | 5 | 4.0 | 5 |

* **SUM**

df.sum()

**OUTPUT**

id 46971

name Kiana LorJoshua LonakerDakota BlancoNatasha Ya...

nationality ChinaUnited States of AmericaUnited States of ...

city SuzhouSanta ClaritaOaklandCastro ValleySão Jos...

latitude 10089.06

longitude -19813.51

gender FMFFFFMMFMFFMMMFFotherMMMMFMMFFMFFFFMMMFMFMFFF...

ethnic.group 0.0

age 6743

english.grade 1034.5

math.grade 1048.2

sciences.grade 1058.1

language.grade 1349.7

portfolio.rating 1224

coverletter.rating 1262.0

refletter.rating 1286

dtype: object

* **MINIMUM**

df.min(

df.min()

**OUTPUT**

id 0

name Aaqil el-Hashemi

nationality Bangladesh

city Aguascalientes

latitude -33.45

longitude -123.13

gender F

ethnic.group NaN

age 19

english.grade 1.5

math.grade 2.1

sciences.grade 1.4

language.grade 1.0

portfolio.rating 1

coverletter.rating 1.0

refletter.rating 1

dtype: object

* **MAXIMUM**

df.max()

**OUTPUT**

id 306

name Zoe Kern

nationality United States of America

city Zülpich

latitude 59.89

longitude 139.75

gender other

ethnic.group NaN

age 26

english.grade 4.0

math.grade 4.0

sciences.grade 4.0

language.grade 5.0

portfolio.rating 5

coverletter.rating 5.0

refletter.rating 5

dtype: object

* **COUNT**

df.count()

**OUTPUT**

id 307

name 307

nationality 307

city 307

latitude 307

longitude 307

gender 307

ethnic.group 0

age 307

english.grade 307

math.grade 307

sciences.grade 307

language.grade 307

portfolio.rating 307

coverletter.rating 307

refletter.rating 307

dtype: int64

* **GROUPBY**

grouped = df.groupby('city')

print(grouped.agg({'age':'sum'}))

**OUTPUT**

age

city

Aguascalientes 40

Alfenas 22

Ankara 21

Baltimore 22

Bangalore 43

... ...

Waukegan 21

West Jordan 22

Wuhan 21

Zamora 23

Zülpich 20

[161 rows x 1 columns]

1. Use Map, Filter, Reduce, and Lambda Functions with Pandas data frames

* MAP

import pandas as pd

# read CSV file into Dataframe

d = pd.read\_csv('/content/student-dataset.csv')

d['english.grade'] = d['english.grade'].map(lambda x: x \* 1.10)

print(d)

* FILTER

filtered\_d = d[d['age'].map(lambda x: x >=30)]

print(filtered\_d)

* REDUCE

from functools import reduce

if not filtered\_d['age'].empty:  # Check if 'age' column is empty

    total = reduce(lambda x, y: x + y, filtered\_d['age'])

    print(total)

else:

    print("The filtered DataFrame is empty, cannot calculate total age.")

OUTPUT

id name nationality city \

0 0 Kiana Lor China Suzhou

1 1 Joshua Lonaker United States of America Santa Clarita

2 2 Dakota Blanco United States of America Oakland

3 3 Natasha Yarusso United States of America Castro Valley

4 4 Brooke Cazares Brazil São José dos Campos

.. ... ... ... ...

302 302 Austin Haas United States of America Columbus

303 303 Madison Fithian United States of America Los Angeles

304 304 Zachary Mulvahill United States of America Los Angeles

305 305 Eliana Michelsen United States of America Oakland

306 306 Dane Whittemore Canada Toronto

latitude longitude gender ethnic.group age english.grade math.grade \

0 31.31 120.62 F NaN 22 8.252817 3.7

1 34.39 -118.54 M NaN 22 6.838048 3.2

2 37.80 -122.27 F NaN 22 9.195996 3.8

3 37.69 -122.09 F NaN 20 7.781227 2.8

4 -23.18 -45.88 F NaN 21 8.724406 2.6

.. ... ... ... ... ... ... ...

302 39.96 -83.00 M NaN 20 8.488612 3.7

303 34.05 -118.24 F NaN 20 8.488612 3.9

304 34.05 -118.24 M NaN 20 7.545433 3.4

305 37.80 -122.27 F NaN 23 7.073843 2.8

306 43.67 -79.42 M NaN 21 8.960201 3.2

sciences.grade language.grade portfolio.rating coverletter.rating \

0 3.1 1.0 4 4.0

1 3.6 5.0 5 4.0

2 3.2 5.0 3 3.0

3 3.2 5.0 5 2.0

4 3.4 1.0 4 4.0

.. ... ... ... ...

302 3.1 5.0 4 5.0

303 4.0 5.0 5 5.0

304 3.9 5.0 5 5.0

305 2.9 5.0 4 4.0

306 4.0 5.0 5 4.0

refletter.rating

0 4

1 5

2 4

3 4

4 5

.. ...

302 5

303 3

304 3

305 5

306 5

[307 rows x 16 columns]

Empty DataFrame

Columns: [id, name, nationality, city, latitude, longitude, gender, ethnic.group, age, english.grade, math.grade, sciences.grade, language.grade, portfolio.rating, coverletter.rating, refletter.rating]

Index: []

The filtered DataFrame is empty, cannot calculate total age

1. Visualize the data set (At least 6 different plots).

import pandas as pd

# load the CSV file into DataFrame

d=pd.read\_csv('/content/student-dataset.csv')

import matplotlib.pyplot as plt

import seaborn as sns

# Set plot size

plt.figure(figsize=(20, 15))

#1. Line plot

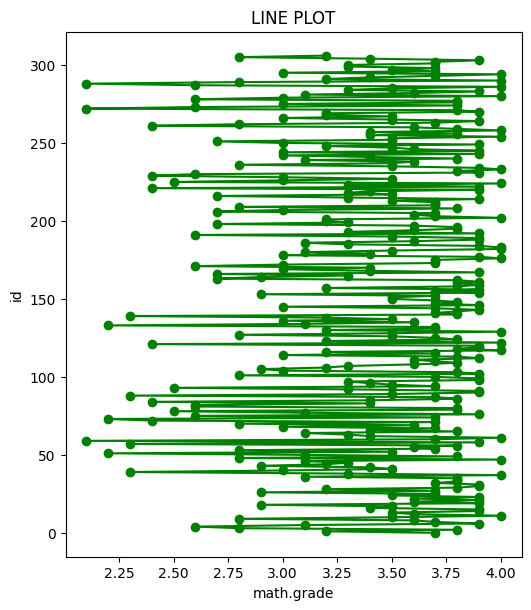
plt.subplot(2, 3, 1)

plt.plot(d['math.grade'], d['id'], marker='o', linestyle='-', color='green')

plt.title("LINE PLOT")

plt.xlabel('math.grade')

plt.ylabel('id')



#2. Bar plot

plt.subplot(1, 1, 1)

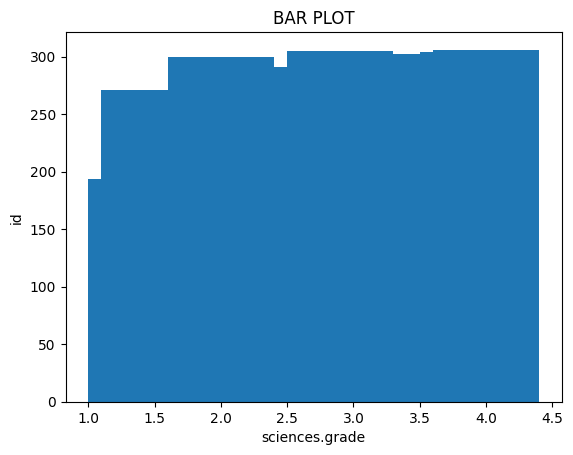
plt.bar(d['sciences.grade'], d['id'])

plt.title("BAR PLOT")

plt.xlabel('sciences.grade')

plt.ylabel('id')

OUTPUT



#3.Histogram plot

plt.subplot(2, 3, 3)

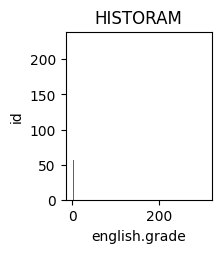
plt.hist(d['english.grade'], d['id'])

plt.title("HISTORAM")

plt.xlabel('english.grade')

plt.ylabel('id')

OUTPUT



#4. Scatter plot

plt.subplot(2, 3, 4)

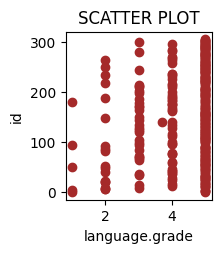
plt.scatter(d['language.grade'], d['id'], marker='o', linestyle='-', color='brown')

plt.title("SCATTER PLOT")

plt.xlabel('language.grade')

plt.ylabel('id')

OUTPUT



#5. Box plot

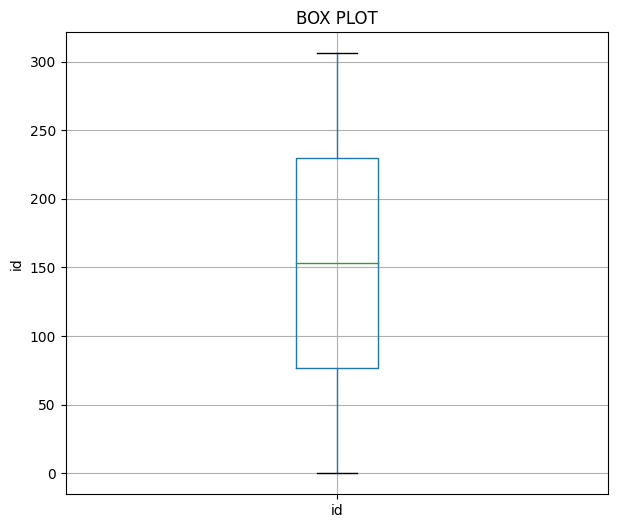
plt.figure(figsize=(7, 6))

d.boxplot(column=['id'])

plt.title("BOX PLOT")

plt.ylabel('id')

OUTPUT



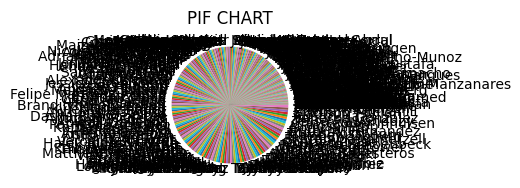
#6. pie chart

plt.subplot(2, 3, 6)

plt.pie(d['id'], labels=d['name'])

plt.title('PIF CHART')

OUTPUT



**DATASET:02**(/content/customers-1001.csv)

**DESCRIPTION**

* Index
* Customer Id
* First Name
* Last Name
* Company
* City
* Country
* Phone 1
* Phone 2
* Email
* Subscription Date
* Website
* Unnamed: 12
* Unnamed: 13
* age

1. **Aggregate functions**

* HEAD

import pandas as pd

path="/content/customers-1001.csv"

df=pd.read\_csv(path)

df.head(5)

OUTPUT

Index Customer Id First Name Last Name Company City Country Phone 1 Phone 2 Email Subscription Date Website Unnamed: 12 Unnamed: 13 age

0 1 DD37Cf93aecA6Dc Sheryl Baxter Rasmussen Group East Leonard Chile 229.077.5154 397.884.0519x718 zunigavanessa@smith.info 24-08-2020 http://www.stephenson.com/ NaN NaN 25

1 2 1Ef7b82A4CAAD10 Preston Lozano Vega-Gentry East Jimmychester Djibouti 5153435776 686-620-1820x944 vmata@colon.com 23-04-2021 http://www.hobbs.com/ NaN NaN 35

2 3 6F94879bDAfE5a6 Roy Berry Murillo-Perry Isabelborough Antigua and Barbuda -1199 (496)978-3969x58947 beckycarr@hogan.com 25-03-2020 http://www.lawrence.com/ NaN NaN 45

3 4 5Cef8BFA16c5e3c Linda Olsen Dominguez, Mcmillan and Donovan Bensonview Dominican Republic 001-808-617-6467x12895 -9892 stanleyblackwell@benson.org 02-06-2020 http://www.good-lyons.com/ NaN NaN 65

4 5 053d585Ab6b3159 Joanna Bender Martin, Lang and Andrade West Priscilla Slovakia (Slovak Republic) 001-234-203-0635x76146 001-199-446-3860x3486 colinalvarado@miles.net 17-04-2021 https://goodwin-ingram.com/ NaN NaN 32

* TAIL

df.tail()

OUTPUT

Index Customer Id First Name Last Name Company City Country Phone 1 Phone 2 Email Subscription Date Website

95 96 cb8E23e48d22Eae Karl Greer Carey LLC East Richard Guyana (188)169-1674x58692 001-841-293-3519x614 hhart@jensen.com 2022-01-30 http://hayes-perez.com/

96 97 CeD220bdAaCfaDf Lynn Atkinson Ware, Burns and Oneal New Bradview Sri Lanka +1-846-706-2218 605.413.3198 vkemp@ferrell.com 2021-07-10 https://novak-allison.com/

97 98 28CDbC0dFe4b1Db Fred Guerra Schmitt-Jones Ortegaland Solomon Islands +1-753-067-8419x7170 +1-632-666-7507x92121 swagner@kane.org 2021-09-18 https://www.ross.com/

98 99 c23d1D9EE8DEB0A Yvonne Farmer Fitzgerald-Harrell Lake Elijahview Aruba (530)311-9786 001-869-452-0943x12424 mccarthystephen@horn-green.biz 2021-08-11 http://watkins.info/

99 100 2354a0E336A91A1 Clarence Haynes Le, Nash and Cross Judymouth Honduras (753)813-6941 783.639.1472 colleen91@faulkner.biz 2020-03-11 <http://www.hatfield-saunders.net/>

* SUM

df.sum()

OUTPUT

Index 5050

Customer Id DD37Cf93aecA6Dc1Ef7b82A4CAAD106F94879bDAfE5a65...

First Name SherylPrestonRoyLindaJoannaAimeeDarrenBrettShe...

Last Name BaxterLozanoBerryOlsenBenderDownsPeckMullenMey...

Company Rasmussen GroupVega-GentryMurillo-PerryDomingu...

City East LeonardEast JimmychesterIsabelboroughBens...

Country ChileDjiboutiAntigua and BarbudaDominican Repu...

Phone 1 229.077.51545153435776+1-539-402-0259001-808-6...

Phone 2 397.884.0519x718686-620-1820x944(496)978-3969x...

Email zunigavanessa@smith.infovmata@colon.combeckyca...

Subscription Date 2020-08-242021-04-232020-03-252020-06-022021-0...

Website [http://www.stephenson.com/http://www.hobbs.com](http://www.stephenson.com/http:/www.hobbs.com)...

dtype: object

* MINIMUM

df.min()

OUTPUT

Index 1

Customer Id 010468dAA11382c

First Name Aimee

Last Name Alvarado

Company Acosta, Petersen and Morrow

City Acevedoville

Country Albania

Phone 1 (041)737-3846

Phone 2 (026)401-7353x2417

Email [aharper@maddox-townsend.org](mailto:aharper@maddox-townsend.org)

Subscription Date 2020-01-13

Website <http://acosta.org/>

dtype: object

\*MAXIMUM

df.max()

OUTPUT

Index 100

Customer Id faCEF517ae7D8eB

First Name Yvonne

Last Name Zuniga

Company Winters-Mendoza

City Zimmermanland

Country Zimbabwe

Phone 1 981-544-9452

Phone 2 999-728-1637

Email [zvalencia@phelps.com](mailto:zvalencia@phelps.com)

Subscription Date 2022-05-26

Website <https://www.vang.com/>

dtype: object

\*COUNT

df.count( )

OUTPUT

Index 100

Customer Id 100

First Name 100

Last Name 100

Company 100

City 100

Country 100

Phone 1 100

Phone 2 100

Email 100

Subscription Date 100

Website 100

dtype: int64

\*GROUPBY

 grouped = df.groupby('Country')

 print(grouped.agg({'Email':'sum'}))

OUTPUT

Email

Country

Albania [mariah88@santos.com](mailto:mariah88@santos.com)

Algeria [xpittman@ritter-carney.net](mailto:xpittman@ritter-carney.net)

Anguilla [dudleyemily@gonzales.biz](mailto:dudleyemily@gonzales.biz)

Antigua and Barbuda [beckycarr@hogan.com](mailto:beckycarr@hogan.com)

Aruba [mccarthystephen@horn-green.biz](mailto:mccarthystephen@horn-green.biz)

... ...

Uzbekistan [bauercrystal@gay.com](mailto:bauercrystal@gay.com)

Vietnam [mark42@robbins.com](mailto:mark42@robbins.com)

Western Sahara [holtgwendolyn@watson-davenport.com](mailto:holtgwendolyn@watson-davenport.com)

Yemen [hilljesse@barrett.info](mailto:hilljesse@barrett.info)

Zimbabwe gfigueroa@boone-zavala.combuckleycory@odonnell...

[85 rows x 1 columns]

1. Use Map, Filter, Reduce, and Lambda Functions with Pandas data frames

\*MAP

import pandas as pd

# read CSV file into Dataframe

df = pd.read\_csv('/content/customers-1001.csv')

df['Index'] = df['Index'].map(lambda x: x \* 1.10)

print(df)

\*FILTER

filtered\_df = df[df['age'].map(lambda x: x >=60)]

print(filtered\_df)

\*REDUCE

total = reduce(lambda x, y: x + y, filtered\_df['Index'])

print(total)

OUTPUT

Index Customer Id First Name Last Name \

0 1.1 DD37Cf93aecA6Dc Sheryl Baxter

1 2.2 1Ef7b82A4CAAD10 Preston Lozano

2 3.3 6F94879bDAfE5a6 Roy Berry

3 4.4 5Cef8BFA16c5e3c Linda Olsen

4 5.5 053d585Ab6b3159 Joanna Bender

.. ... ... ... ...

95 105.6 cb8E23e48d22Eae Karl Greer

96 106.7 CeD220bdAaCfaDf Lynn Atkinson

97 107.8 28CDbC0dFe4b1Db Fred Guerra

98 108.9 c23d1D9EE8DEB0A Yvonne Farmer

99 110.0 2354a0E336A91A1 Clarence Haynes

Company City \

0 Rasmussen Group East Leonard

1 Vega-Gentry East Jimmychester

2 Murillo-Perry Isabelborough

3 Dominguez, Mcmillan and Donovan Bensonview

4 Martin, Lang and Andrade West Priscilla

.. ... ...

95 Carey LLC East Richard

96 Ware, Burns and Oneal New Bradview

97 Schmitt-Jones Ortegaland

98 Fitzgerald-Harrell Lake Elijahview

99 Le, Nash and Cross Judymouth

Country Phone 1 \

0 Chile 229.077.5154

1 Djibouti 5153435776

2 Antigua and Barbuda -1199

3 Dominican Republic 001-808-617-6467x12895

4 Slovakia (Slovak Republic) 001-234-203-0635x76146

.. ... ...

95 Guyana (188)169-1674x58692

96 Sri Lanka -3769

97 Solomon Islands +1-753-067-8419x7170

98 Aruba (530)311-9786

99 Honduras (753)813-6941

Phone 2 Email Subscription Date \

0 397.884.0519x718 [zunigavanessa@smith.info](mailto:zunigavanessa@smith.info) 24-08-2020

1 686-620-1820x944 [vmata@colon.com](mailto:vmata@colon.com) 23-04-2021

2 (496)978-3969x58947 [beckycarr@hogan.com](mailto:beckycarr@hogan.com) 25-03-2020

3 -9892 [stanleyblackwell@benson.org](mailto:stanleyblackwell@benson.org) 02-06-2020

4 001-199-446-3860x3486 [colinalvarado@miles.net](mailto:colinalvarado@miles.net) 17-04-2021

.. ... ... ...

95 001-841-293-3519x614 [hhart@jensen.com](mailto:hhart@jensen.com) 30-01-2022

96 605.413.3198 [vkemp@ferrell.com](mailto:vkemp@ferrell.com) 10-07-2021

97 +1-632-666-7507x92121 [swagner@kane.org](mailto:swagner@kane.org) 18-09-2021

98 001-869-452-0943x12424 [mccarthystephen@horn-green.biz](mailto:mccarthystephen@horn-green.biz) 11-08-2021

99 783.639.1472 [colleen91@faulkner.biz](mailto:colleen91@faulkner.biz) 11-03-2020

Website Unnamed: 12 Unnamed: 13 age

0 <http://www.stephenson.com/> NaN NaN 25

1 <http://www.hobbs.com/> NaN NaN 35

2 <http://www.lawrence.com/> NaN NaN 45

3 <http://www.good-lyons.com/> NaN NaN 65

4 <https://goodwin-ingram.com/> NaN NaN 32

.. ... ... ... ...

95 <http://hayes-perez.com/> NaN NaN 39

96 <https://novak-allison.com/> NaN NaN 49

97 <https://www.ross.com/> NaN NaN 59

98 <http://watkins.info/> NaN NaN 69

99 <http://www.hatfield-saunders.net/> NaN NaN 25

[100 rows x 15 columns]

Index Customer Id First Name Last Name \

3 4.4 5Cef8BFA16c5e3c Linda Olsen

9 11.0 8C2811a503C7c5a Michelle Gallagher

16 18.7 a5DC21AE3a21eaA Caroline Foley

22 25.3 2FB0FAA1d429421 Colleen Howard

38 42.9 dca4f1D0A0fc5c9 Bruce Esparza

43 48.4 D3fC11A9C235Dc6 Luis Greer

45 50.6 fD780ED8dbEae7B Joanne Montes

47 52.8 283DFCD0Dba40aF Gloria Mccall

56 62.7 c8FE57cBBdCDcb2 Phyllis Maldonado

71 79.2 Ef859092FbEcC07 Richard Roth

72 80.3 F560f2d3cDFb618 Candice Keller

82 91.3 E8E7e8Cfe516ef0 Hunter Moreno

98 108.9 c23d1D9EE8DEB0A Yvonne Farmer

Company City Country \

3 Dominguez, Mcmillan and Donovan Bensonview Dominican Republic

9 Beck-Hendrix Elaineberg Timor-Leste

16 Winters-Mendoza West Adriennestad Western Sahara

22 Greer and Sons Brittanyview Paraguay

38 Huerta-Mclean Poolefurt Montenegro

43 Cross PLC North Drew Bulgaria

45 Price, Sexton and Mcdaniel Gwendolynview Palau

47 Brennan, Acosta and Ramos North Kerriton Ghana

56 Costa PLC Lake Whitney Saint Barthelemy

71 Conway-Mcbride New Jasmineshire Morocco

72 Huynh and Sons East Summerstad Zimbabwe

82 Fitzpatrick-Lawrence East Clinton Isle of Man

98 Fitzgerald-Harrell Lake Elijahview Aruba

Phone 1 Phone 2 \

3 001-808-617-6467x12895 -9892

9 739.218.2516x459 001-054-401-0347x617

16 936.222.4746x9924 001-469-948-6341x359

22 1935085151 (947)115-7711x5488

38 559-529-4424 001-625-000-7132x0367

43 001-336-025-6849x701 684.698.2911x6092

45 (897)726-7952 (467)886-9467x5721

47 445-603-6729 001-395-959-4736x4524

56 4500370767 001-508-064-6725x017

71 581-440-6539 9857827463

72 001-927-965-8550x92406 001-243-038-4271x53076

82 (733)833-6754 001-761-013-7121

98 (530)311-9786 001-869-452-0943x12424

Email Subscription Date \

3 [stanleyblackwell@benson.org](mailto:stanleyblackwell@benson.org) 02-06-2020

9 [mdyer@escobar.net](mailto:mdyer@escobar.net) 08-11-2021

16 [holtgwendolyn@watson-davenport.com](mailto:holtgwendolyn@watson-davenport.com) 10-03-2021

22 [rsingleton@ryan-cherry.com](mailto:rsingleton@ryan-cherry.com) 19-08-2020

38 [preese@frye-vega.com](mailto:preese@frye-vega.com) 22-10-2021

43 [bstuart@williamson-mcclure.com](mailto:bstuart@williamson-mcclure.com) 15-05-2022

45 [juan80@henson.net](mailto:juan80@henson.net) 01-07-2020

47 [bartlettjenna@zuniga-moss.biz](mailto:bartlettjenna@zuniga-moss.biz) 11-03-2022

56 [yhanson@warner-diaz.org](mailto:yhanson@warner-diaz.org) 25-01-2021

71 [aharper@maddox-townsend.org](mailto:aharper@maddox-townsend.org) 23-02-2020

72 [buckleycory@odonnell.net](mailto:buckleycory@odonnell.net) 22-08-2020

82 [isaac26@benton-finley.com](mailto:isaac26@benton-finley.com) 28-12-2020

98 [mccarthystephen@horn-green.biz](mailto:mccarthystephen@horn-green.biz) 11-08-2021

Website Unnamed: 12 Unnamed: 13 age

3 <http://www.good-lyons.com/> NaN NaN 65

9 <https://arias.com/> NaN NaN 62

16 <http://www.benson-roth.com/> NaN NaN 66

22 <http://paul.biz/> NaN NaN 63

38 <http://www.farley.org/> NaN NaN 69

43 <https://fletcher-nielsen.com/> NaN NaN 71

45 <http://ochoa.com/> NaN NaN 75

47 <http://burgess-frank.com/> NaN NaN 63

56 <http://www.bernard.com/> NaN NaN 66

71 <https://www.brooks.com/> NaN NaN 95

72 <https://www.lucero.com/> NaN NaN 63

82 <http://walls.info/> NaN NaN 66

98 <http://watkins.info/> NaN NaN 69

676.5

(c) Visualize the data set (At least 6 different plots).

import pandas as pd

# load the CSV file into DataFrame

df=pd.read\_csv('/content/customers-1001.csv')

import matplotlib.pyplot as plt

#1. Line plot

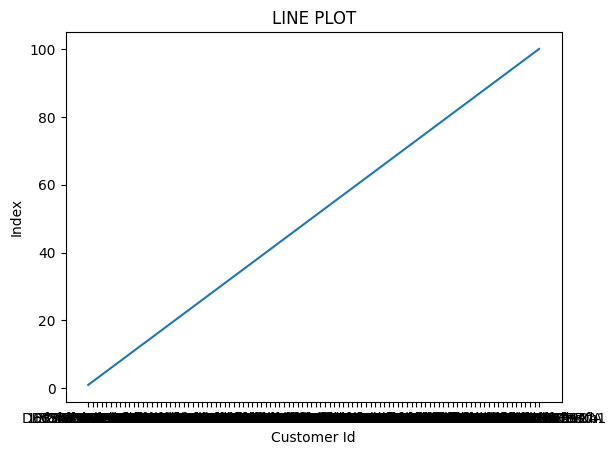
plt.plot(df['Customer Id'], df['Index'], )

plt.title("LINE PLOT")

plt.xlabel('Customer Id')

plt.ylabel('Index')

OUTPUT



#2. Bar plot

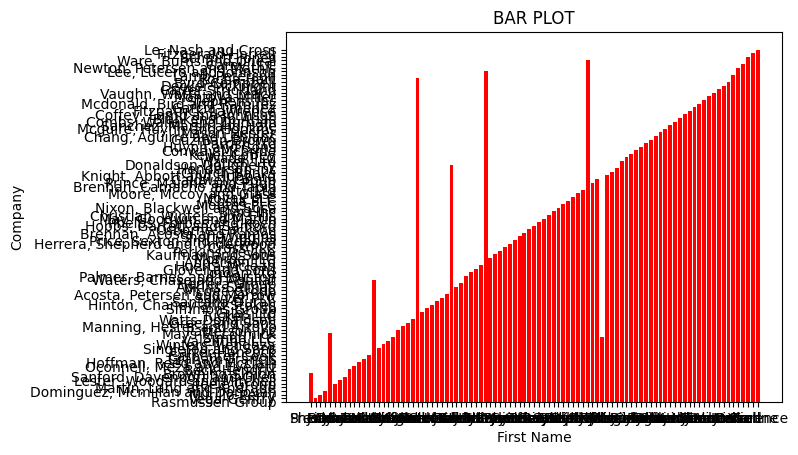
plt.bar(df['First Name'], df['Company'], color="red")

plt.title("BAR PLOT")

plt.xlabel('First Name')

plt.ylabel('Company')

OUTPUT



#3.Histogram plot

plt.hist(df['Last Name'], df['Index'], color="brown")

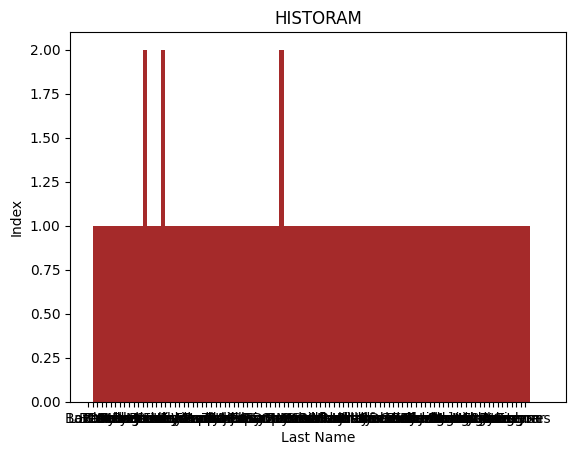
plt.title("SCATTER PLOT")

plt.title("HISTORAM")

plt.xlabel('Last Name')

plt.ylabel('Index')

OUTPUT



#4. Scatter plot

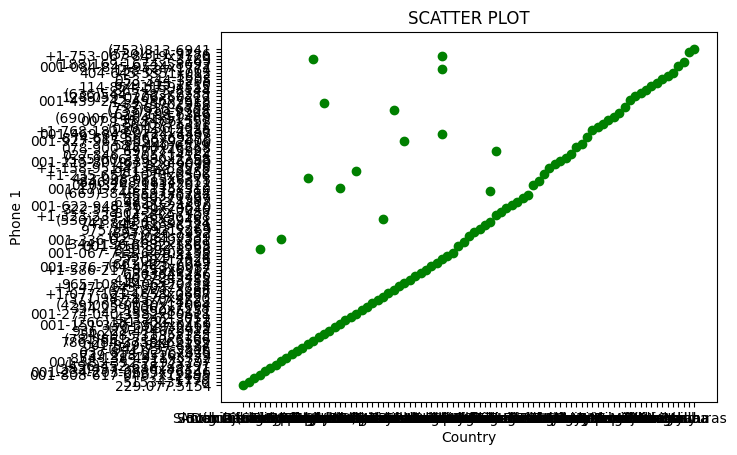
plt.scatter(df['Country'], df['Phone 1'], marker='o', linestyle='-', color='green')

plt.title("SCATTER PLOT")

plt.xlabel('Country')

plt.ylabel('Phone 1')

OUTPUT



#5. Box plot

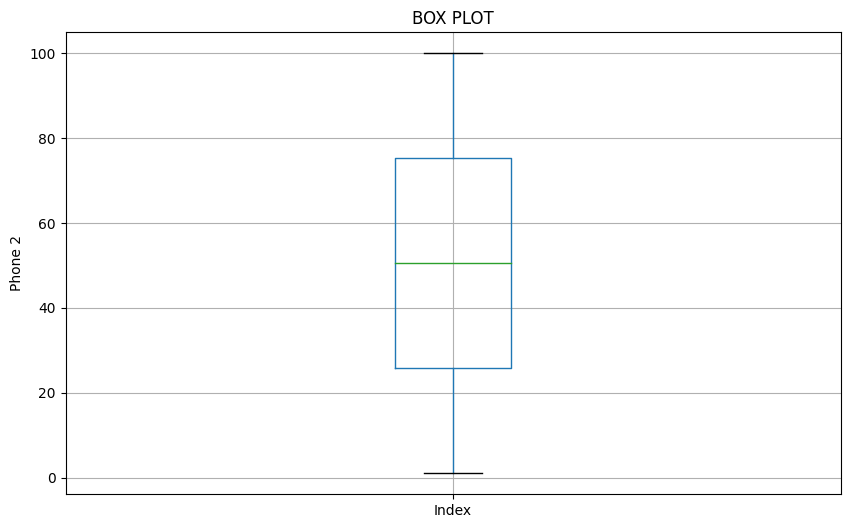
plt.figure(figsize=(10, 6))

df.boxplot(column=['Index'])

plt.title("BOX PLOT")

plt.ylabel('Phone 2')

OUTPUT



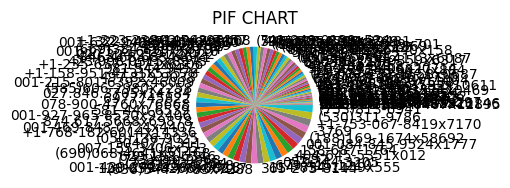
#6. pie chart

plt.subplot(2, 3, 6)

plt.pie(df['Index'], labels=df['Phone 1'])

plt.title('PIF CHART')

OUTPUT



**(D) How do you create a project plan and product backlog for an Al project? (Everyone chooses the area you want to work on or do the research work. Give a brief introduction to the work carried out and the final report to be submitted at the end of the course.)**

**a. Create a Git Repository for following the Regression Project ML/deep learning.**

**b. Classification Project-ML/deep learning**

**c. Clustering project-ML/deep learning**

**d. Natural Language Processing-ML/deep learning**

Here's a breakdown of how to create project plans and product backlogs for AI projects, along with guidance on setting up Git repositories for each type of project:

\*Project Planning and Backlog Creation\*

1. Define the Project Scope\*

\* \*Project Goal:\* Clearly state the overall objective of the AI project. What problem are you trying to solve? What are the desired outcomes?

\* \*Target Audience:\* Who will benefit from this project?

\* \*Data Requirements:\* What kind of data will you need? Where will you get it? What are the data quality expectations?

\* \*Technical Constraints:\* Are there any limitations on the technology you can use (e.g., specific libraries, hardware)?

2. Create a Project Plan\*

\* \*Project Timeline:\* Break down the project into manageable phases with estimated durations.

\* \*Milestones:\* Define key deliverables and deadlines.

\* \*Resources:\* Identify the team members, tools, and software needed.

\* \*Risk Assessment:\* Identify potential challenges and develop mitigation strategies.

3. Develop a Product Backlog\*

\* \*User Stories:\* Write user stories that describe the features and functionality of the AI system from the user's perspective.

\* \*Prioritization:\* Rank user stories based on their importance and value.

\* \*Acceptance Criteria:\* Define clear criteria for determining when a user story is complete.

\*Example Project Plan and Backlog (Regression Project)\*

\*Project Goal:\* Develop a machine learning model to predict house prices based on historical data.

\*Project Timeline:\*

\* \*Phase 1 (2 weeks):\* Data Collection and Preparation

\* \*Phase 2 (2 weeks):\* Model Selection and Training

\* \*Phase 3 (1 week):\* Model Evaluation and Optimization

\* \*Phase 4 (1 week):\* Deployment and Documentation

\*Product Backlog:\*

\* \*User Story 1:\* As a real estate agent, I want to be able to predict house prices based on features like size, location, and amenities so that I can provide accurate valuations to clients.

\* \*Acceptance Criteria:\* The model should achieve an R-squared score of at least 0.8 on the test set.

\* \*User Story 2:\* As a homeowner, I want to be able to understand the factors that influence house prices so that I can make informed decisions about selling or renovating my property.

\* \*Acceptance Criteria:\* The model should provide insights into the relative