

NAME: *SHOAIB SAOUD*

DATABASE SYSTEMS

COURSEWORK

TASK 1:

- A)** The purpose of converting a given dataset into a relational database is to organize the data in a structured and systematic manner that can be easily searched for and accessed. There should be multiple tables in this dataset that contain various types of information, including the Owner table, which will contain details regarding the owner. Dog tables, grooming tables, booking tables, etc. also seem to be similar. Additionally, the provided data consists of some data anomalies like insertion anomalies, deletion anomalies, and modification anomalies.

INSERTION ANOMALIES:

Add a new owner to the dataset; repeat the grooming instructions. Additionally, if the dog receives two or more treatments, the owner's information must be repeated several times.

DELETION ANOMALY:

Deleting the owner "Wen Chung" from the dataset will also remove the existence of the Groom Option ("C011" or "Lion Cut") from this dataset. Similarly, The Groom option ("N055 or "Nail Trim & Paint") will no longer be present in this dataset if the owner "Colin Daniels" is deleted from it.

MODIFICATION ANOMALY:

Each dog requires a different set of treatments, which must be repeated along with the dog's details, as well as different types of grooming options.

ADVANTAGES:

- Provides a straightforward way to organize data, which makes it easier to manage and analyze using various tables, such as owner, dog, groom, and booking tables, etc.
- Using SQL queries and commands to quickly retrieve data to obtain the data required for the particular purpose.
- Using foreign keys, each table can be connected to another with the same values.
- Able to use normalization and SQL queries to manipulate large amounts of data.

B)

Demonstrate the detailed methods of implementing the relational database:

i- Normalization stage and Justification



normalization
process.xlsx

1NF

The information about the Owner, Dog, Groom, and Booking are all separated in this table. The "Owner ID" is the primary key in the Owner table because it uniquely identifies the Owner. The table also contains the Owner's Email Address and Phone Number. as well as their Name. The primary key in the Dog Table is "Dog ID"; the table also includes "Dog Name", "Comments", and physical characteristics of the

dog. The "Owner ID" is used to identify that table as a foreign table. The "Groom Option", which identifies the type of Grooming Service, is the primary key in the Groom Table, as well as the Table also includes a "Description", "Groom Cost", and "Groom Time" of the Treatment. The attributes in the booking table include the "Booking date" and "Booking time", the "Total Length of Appointment", and the "TOTAL Cost". The booking table also lists the following foreign keys: "Dog ID", "Owner ID", and "Groom Option". As you can see, this table does not contain a primary key.

2NF

The attributes "Owner Name", "Tel No", and "Email" are completely dependent on "Owner ID", as can be seen in the owner table. Dog table attributes like "Dog Name", "Breed", and so forth are entirely dependent on Dog ID. Grooming details such as "Description", "Groom cost", and "Groom time" are entirely dependent on Groom Option. "Groom option", "Description", and "Booking date" are entirely dependent on Dog ID. "Owner ID", "Booking Time", "Total Length of Appointment", and "TOTAL Cost", On the other hand, are not only dependent on "Groom Option" but are completely dependent on "Dog ID".

3NF

Because there is no transitive dependency, the information in the Owner Table will remain unchanged or will already be in the 3NF. In the Dog Table the "Owner ID" column is functional dependent on the Dog ID column, we would need to separate it into a new table with the "Dog ID" (foreign key) and "Dog Name". Moreover, each dog entry has a unique identity (Dog ID) as the primary key, and all non-primary key columns ("Breed", "Size", "Coat", and "Comments") are directly dependent on the primary key. There are no transitive dependencies, therefore the grooming table will remain the same or is already in the 3NF. The dog-grooming table will remain unchanged. The booking tables, which comprise the "Booking date", "Booking time", "Total Length of Appointment", "TOTAL Cost". and other attributes ("Owner ID", "Dog ID"), will also remain the same.

ii- Physical” Entity-Relationship Diagram using Crow’s Foot notation and Explanation.



ER diagram.drawio (Command Line)

In the ER-diagram:

The owner has a zero or one notation on one side of the relationship and a one through many notations on the other side with the dog since a single owner can have several owners and a dog can have no owner. A dog has a one through many notations on one side of a relationship and a one and only one on the other with the breed because a dog can have only one kind of breed or multiple dogs can have the same kind of breed. Dogs have a one through many relationships with grooming on both sides. Because many dogs or a single dog may require various grooming treatments and vice versa. Groom has a one through many relationships with the booking on both sides of the relationship as various grooming treatments might well have numerous bookings and vice versa. Dog has a one through many relationships with the booking on both ends since a dog may have several booking appointments or vice-versa.

iii- SQL commands to create the tables and screenshots

CREATE TABLE owner (

Owner ID INTEGER PRIMARY KEY

NOT NULL,

Owner Name CHAR (50),

Tel No INTEGER,

Email VARCHAR (50)

);

	OwnerID	OwnerName	TelNo	Email
1	101	Joe Brown	0121 543182	Jbrown@gmaily.com
2	212	Samantha Kibson	7890 67123	kibsons@yakkity.co.uk
3	213	Jordan Foster	6734 321022	FosterMJ@Yahoop.com
4	231	Wen Chung	533 789873	wchung@btbt.com
5	398	Penny Patel	7684 298871	ppatel@gsnail.com
6	452	Colin Daniels	2476 90345	cs@aool.com
7	542	Peter Small	123 654263	small@gsnail.com
8	620	Fathima Masud	342 907652	Masud@gsnail.com
9	765	Mahima Raab	543 777123	raabm@btd.com
10	776	Samuel Johnson	231 592833	Johnsons@yakkity.com

CREATE TABLE dog (

Dog ID VARCHAR (5) PRIMARY KEY

NOT NULL,

Breed CHAR (50) NOT NULL,

REFERENCES dog (Breed),

Size CHAR (10) NOT NULL,

Coat CHAR (10) NOT NULL,

Comments CHAR (200)

);

	DogID	DogName	Breed	Size	Coat	Comments
1	D901	Bella	Rhodesian Ridgeback	Large	Single	Cannot eat beef based treats
2	D321	Poppy-Sage	Pomeranian	Small	Double	Needs a muzzle for nail trim
3	D877	Rick-Roly	Schnauzer	Small	Single	Use sensitive shampoo
4	D567	Brian	Dachshund	Small	Single	Snappy when being washed
5	D231	Max	Golden Retriever	Large	Double	Does not like hairdryer on face
6	D987	Diesel	Dalmatian	Medium	Single	NULL
7	D543	Major Tom	Tibetan Terrier	Medium	Double	NULL
8	D696	Bitsy-Malone	Toy Poodle	Small	Single	Nervous with other dogs
9	D121	Stan	Standard Poodle	Medium	Single	Does not like paws being clipped
10	D887	Dumbledore	Akita	Medium	Double	Sensitive Ears

```

CREATE TABLE dog owner (
    Owner ID INT NOT NULL,
    Dog ID VARCHAR (5) NOT NULL,
    Dog Name CHAR (20) NOT NULL,
    FOREIGN KEY (
        Owner ID
    )
    REFERENCES owner (Owner ID),
    FOREIGN KEY (
        Dog ID
    )
    REFERENCES dog ( Dog ID)
);

```

	OwnerID*	DogID*	DogName
1	765	D901	Bella
2	212	D321	Poppy-Sage
3	542	D877	Rick-Roly
4	213	D567	Brian
5	101	D231	Max
6	620	D987	Diesel
7	776	D543	Major Tom
8	452	D696	Bitsy-Malone
9	231	D121	Stan
10	398	D887	Dumbledore

```

CREATE TABLE groom (
    Groom Option VARCHAR (5) PRIMARY KEY
        NOT NULL,
    Description CHAR (50) NOT NULL,
    Groom Time INT NOT NULL,
    Groom Cost DECIMAL(10,2) NOT NULL
);

```

	GroomOption	Description	GroomTime	GroomCost
1	W011	Wash and Dry	15	£10.00
2	E899	Ear Cleaning	10	£7.50
3	N045	Nail Trim	10	£5.00
4	W050	Luxury Wash and Dry	25	£17.50
5	C021	Teddy Bear Cut	25	£30.00
6	C034	Breed Trim	15	£20.00
7	X110	Undercoat Reduction	10	£7.00
8	N055	Nail Trim & Paint	20	£10.00
9	C011	Lion Cut	25	£30.00

```

CREATE TABLE dog groom (
    Dog ID    VARCHAR (5) NOT NULL,
    Groom Option VARCHAR (5) NOT NULL,
    Description CHAR (100) NOT NULL,
    Booking Date VARCHAR(20) NOT NULL,
    FOREIGN KEY (
        Dog ID
    )
    REFERENCES dog (Dog ID),
    FOREIGN KEY (
        Groom Option
    )
    REFERENCES groom (Groom Option)
);

```


	DogID*	GroomOption*	Description	BookingDate
1	D901	W011	Wash and Dry	8/10/2022
2	D901	E899	Ear Cleaning	8/10/2022
3	D321	N045	Nail Trim	12/16/2022
4	D321	W050	Luxury Wash an...	12/16/2022
5	D321	C021	Teddy Bear Cut	12/16/2022
6	D877	W011	Wash and Dry	4/5/2022
7	D877	C034	Breed Trim	4/5/2022
8	D567	N045	Nail Trim	2/27/2022
9	D567	E899	Ear Cleaning	2/27/2022
10	D231	W011	Wash and Dry	8/10/2022
11	D231	C034	Breed Trim	8/10/2022
12	D231	X110	Undercoat ...	8/10/2022
13	D987	N045	Nail Trim	2/26/2022
14	D987	W011	Wash and Dry	2/26/2022
15	D321	N045	Nail Trim	4/5/2022
16	D901	W011	Wash and Dry	7/12/2022
17	D901	E899	Ear Cleaning	7/12/2022
18	D901	N045	Nail Trim	7/12/2022
19	D321	N045	Nail Trim	6/3/2022
20	D321	W050	Luxury Wash an...	6/3/2022
21	D321	C021	Teddy Bear Cut	6/3/2022
22	D321	X110	Undercoat ...	6/3/2022
23	D901	W011	Wash and Dry	2/11/2022
24	D901	E899	Ear Cleaning	2/11/2022

CREATE TABLE booking (

Dog ID VARCHAR (5) NOT NULL,

Owner ID INT NOT NULL,

Booking Date VARCHAR (20) NOT NULL,

Booking Time VARCHAR (20) NOT NULL,

Total Length of Appointment,

TOTAL Cost DECIMAL (10, 2) NOT NULL,

FOREIGN KEY (

Owner ID

)

REFERENCES owner (Owner ID),

FOREIGN KEY (

Dog ID

)

REFERENCES dog (Dog ID)

);

	OwnerID	DogID*	BookingDate	BookingTime	TOTAL	TOTALCOST
1	765	D901	2/11/2022	14:00	47	£30.50
2	620	D987	2/26/2022	10:00	31	£27.50
3	452	D696	2/26/2022	11:30	70	£57.50
4	213	D567	2/27/2022	9:45	20	£12.50
5	231	D121	3/31/2022	16:00	60	£60.00
6	212	D321	4/5/2022	11:00	10	£5.00
7	542	D877	4/5/2022	14:30	30	£30.00
8	213	D567	6/3/2022	11:30	45	£30.00
9	212	D321	6/3/2022	15:45	70	£64.50
10	776	D543	7/12/2022	9:30	37.5	£26.25
11	765	D901	7/12/2022	11:00	47	£30.50
12	398	D887	7/12/2022	14:15	70	£61.25
13	101	D231	8/10/2022	9:00	60	£55.50
14	765	D901	8/10/2022	15:00	37	£25.50
15	101	D231	10/15/2022	9:15	60	£55.50
16	212	D321	10/15/2022	15:00	60	£52.50
17	212	D321	12/16/2022	11:00	60	£52.50

c)

SQL code for the queries together with a screenshot and Explanation

i-

select Owner Name as ' Owner Name' ,

Tel No as 'Telephone Number',

Email as 'Email Address',

Dog Name as 'Dog Name'

from owner , dog owner

where owner. Owner ID =dog owner." Owner ID*"

	Onwer Name	Telephone Number	Email Address	Dog Name
1	Mahima Raab	543777123	raabm@btd.com	Bella
2	Samantha Kibson	789067123	kibsons@yakkity.co.uk	Poppy-Sage
3	Peter Small	123654263	small@gsnail.com	Rick-Roly
4	Jordan Foster	6734321022	FosterMJ@Yahoop.com	Brian
5	Joe Brown	121543182	Jbrown@gmai.y.com	Max
6	Fathima Masud	342907652	Masud@gsnail.com	Diesel
7	Samuel Johnson	231592833	Johnsons@yakkity.com	Major Tom
8	Colin Daniels	247690345	cs@aool.com	Bitsy-Malone
9	Wen Chung	533789873	wchung@btbt.com	Stan
10	Penny Patel	7684298871	ppatel@gsnail.com	Dumbledore

The SQL code above 'SELECT' the owner's name AS 'Owner Name', Tel no AS 'Telephone Number',

email AS 'Email Address' and dog name AS 'Dog Name' 'FROM' two tables: owner and dog owner. The query specifies a condition in the 'WHERE' clause to join the two tables by matching the "Owner ID" column in the owner table to the "Owner ID" column in the dog table.

ii-

```

select owner. Owner Name, Tel No, Email, booking. "Dog ID*",count(*) as
'Number Of Visits'

from booking ,owner

where owner. Owner ID = booking . "Owner ID*"

group by Booking. "Dog ID*"

order by Number of Visits desc

```

	OwnerName	TelNo	Email	DogID*	NumberOfVisits
1	Samantha Kibson	7890 67123	kibsons@yakkity.co.uk	D321	4
2	Mahima Raab	543 777123	raabm@btd.com	D901	3
3	Jordan Foster	6734 321022	FosterMJ@Yahoop.com	D567	2
4	Joe Brown	0121 543182	Jbrown@gmaily.com	D231	2
5	Fathima Masud	342 907652	Masud@gsnail.com	D987	1
6	Penny Patel	7684 298871	ppatel@gsnail.com	D887	1
7	Peter Small	123 654263	small@gsnail.com	D877	1
8	Colin Daniels	2476 90345	cs@aool.com	D696	1
9	Samuel Johnson	231 592833	Johnsons@yakkity.com	D543	1
10	Wen Chung	533 789873	wchung@btbt.com	D121	1

The SQL query above 'SELECT' the owner's name, telephone number, email, and Dog ID 'FROM' two tables: owner and booking. and 'COUNT' Number of Visits. By matching the "Owner ID" column in the owner table with the " Owner ID" column in the Booking table, the query defines a condition in the "WHERE" clause to link the two data. The query then 'GROUP BY' the results by Dog ID and 'ORDER BY' them in descending order based on the Number of Visits.

iii-

```

select Owner Name, sum(Total Cost) as 'Total Spent'

from owner, booking

where owner. Owner ID = booking. "Owner ID*"

group by owner name

order by TOTAL Cost asc

limit 1

```

	ownername	Total Spent
1	Samantha Kibson	174.5

The SQL code above 'SELECT' the owner's name and the TOTAL Cost, FROM two tables: owner and booking. By matching the "Owner ID" column in the owner table with the "Owner ID" column in the Booking table, the query defines a condition in the "WHERE" clause to link the two data. The query calculates the SUM of the Total Cost AS 'TOTAL SPENT' column for each owner, 'GROUP BY' the result set by owner name, 'ORDER BY' the result set by the TOTAL SPENT in ascending order, and 'LIMIT' the result set to only the first row.

iv-

select Description as 'Treatment Name', COUNT(*) as 'Number Of Requests'

from dog groom

group by Description

order by Number Of Requests DESC

	TreatmentName	NumberOfRequests
1	Nail Trim	10
2	Wash and Dry	8
3	Luxury Wash and Dry	7
4	Ear Cleaning	5
5	Teddy Bear Cut	4
6	Breed Trim	4
7	Undercoat Reduction	3
8	Nail Trim & Paint	1
9	Lion Cut	1

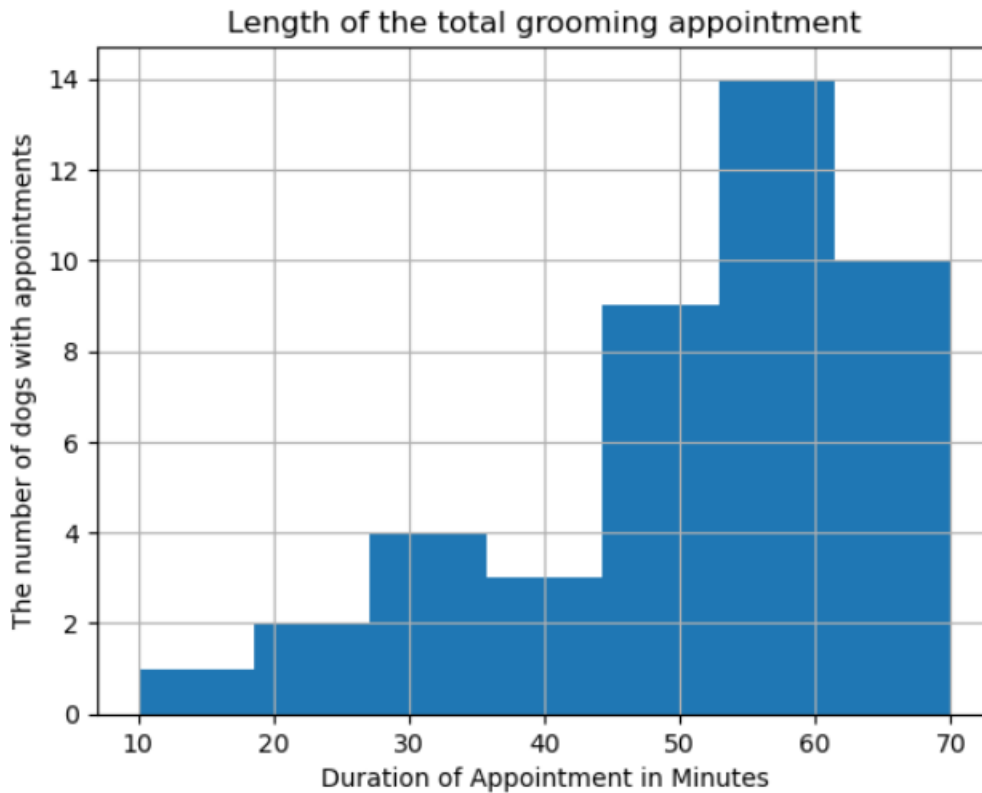
The SQL query above 'SELECT' the Description 'AS' 'Treatment Name ' column 'FROM' the dog groom table and 'COUNT' the number of times each treatment has been requested 'AS' 'Number Of Requests'. The result is 'GROUP BY' Description and 'ORDER BY' the Number Of Requests column in descending order.

TASK 2

Data Visualization

FOR HISTOGRAM

```
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
data = pd.read_csv(r"C:\Users\shoai\Desktop\MutzNutzBookingSystem.csv")
data.hist("Total Length of Appointment", bins=7)
plt.title("Length of the total grooming appointment")
plt.xlabel("Duration of Appointment in Minutes")
plt.ylabel("The number of dogs with appointments")
```



The Total Length of Appointment variable for dog grooming indicates the total duration for each appointment. The purpose of using Total Length of Appointment in a histogram is to understand the distribution of appointment lengths throughout a dataset. The histogram illustrates the frequency distribution of appointment lengths graphically. The appointment durations in minutes are represented on the x-axis, and the frequency (Number of dogs with appointments) of appointments that occur within each length interval or bin is represented on the y-axis. From the Histogram(graph), we can see that the range of appointment lengths varies from 10 minutes to 70 minutes, with some values repeated more frequently than others. The histogram also shows that the most common appointment length is 60 minutes, with a frequency of 14, followed by 70 minutes, with a frequency of 10 etc. Additionally, there are also noticeable peaks around 45 minutes and 47 minutes, which suggests that this is a popular length for appointments. Overall, the histogram is a useful way to summarize the distribution of appointment lengths and identifying trends in data. It also demonstrates that most appointments are one hour in duration, but there are some shorter and longer sessions as well.

FOR BAR CHART

```
import pandas as pd
```

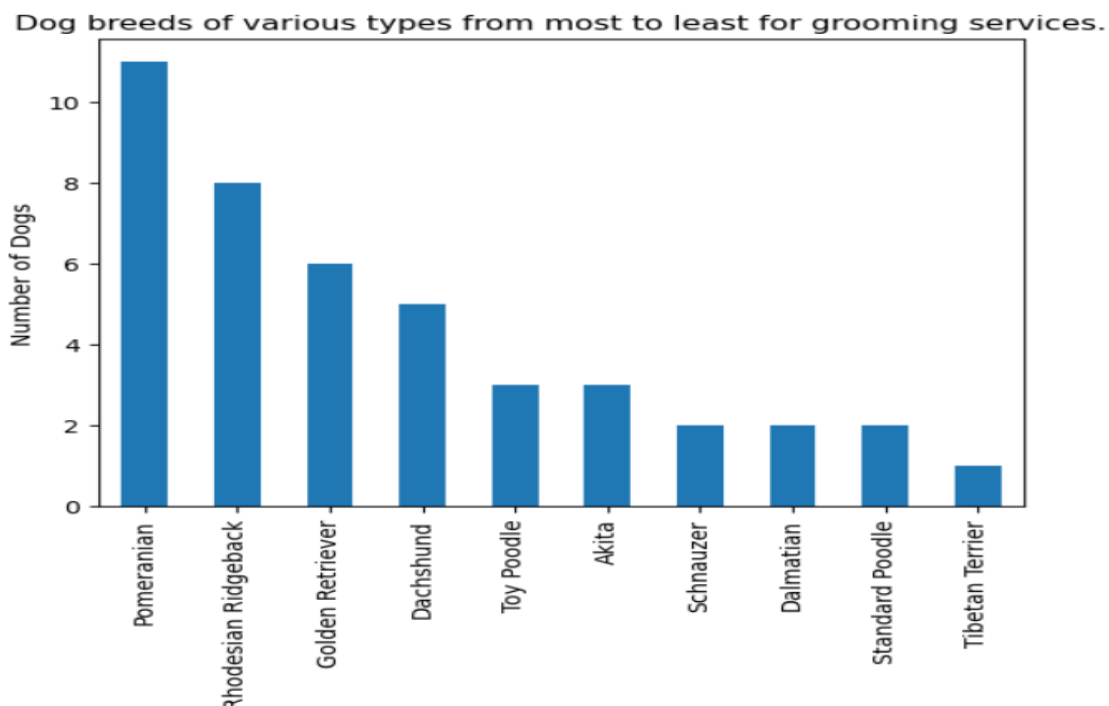
```
import matplotlib.pyplot as plt
```

```
data = pd.read_csv(r"C:\Users\shoai\Desktop\MutzNutzBookingSystem.csv")
```

```
data["Breed"].value_counts().plot(kind="bar")
```

```
plt.title("Dog breeds of various types from most to least for grooming services.")
```

```
plt.ylabel("Number of Dog")
```



The bar chart can be used to analyze patterns or trends in the data and can assist visualize the relative popularity of each dog breed that attends the grooming service. It gives a straightforward and easy-to-read means of comparing the frequency of various dog breeds, and we can illustrate the frequency of each breed using a bar chart. The x-axis of the bar chart indicates each dog breed, while the y-axis represents the frequency or count of each breed that has visited for the grooming service. Each breed will have its own vertical bar displaying its frequency. The graph shows that the most common dog breeds that attended for grooming were Pomeranian and Rhodesian Ridgeback, with a frequency of 11 and 8,

respectively. Golden retrievers and Dachshunds are the next most popular breeds, with each receiving more than four visits. Other breeds, such as Toy Poodles, Akitas, Schnauzers, Dalmatians, Standard Poodles, and Tibetan Terriers, have a lower frequency of visits than the top four. The use of a bar chart allows you to compare the number of visits for each type of dog in a visual and easy-to-understand manner. The bars can be simply compared and arranged depending on their heights, making it simple to determine which breed is the most popular and least popular among visitors. Furthermore, the chart is straightforward to understand, especially for those who are unfamiliar with data visualization.

FOR PIE CHART

```
import pandas as pd
```

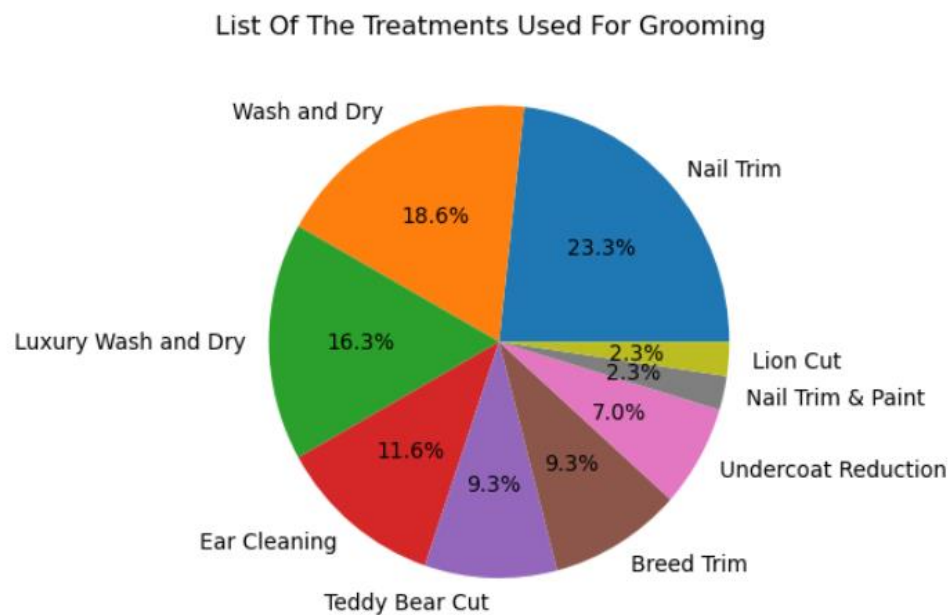
```
import matplotlib.pyplot as plt
```

```
data = pd.read_csv(r"C:\Users\shoai\Desktop\MutzNutzBookingSystem.csv")
```

```
data["Description"].value_counts().plot(kind="pie",autopct='%1.1f%%')
```

```
plt.title("List Of The Treatments Used For Grooming ")
```

```
plt.ylabel("")
```



The pie chart would represent the frequency or percentage of each grooming treatment that was provided. The chart comprises nine sectors (each indicating a grooming session) with various colors to differentiate between the treatments. which also help identify the most popular and the least popular services. According to the pie chart, the most popular grooming treatment is "Nail Trim," which accounts for about 23% of all treatments performed. The next most popular treatments are "Wash and Dry" and "Luxury Wash and Dry," which account for approximately 18% and 16% of all treatments, respectively. Following that is "Ear Cleansing," which has a proportion of roughly 11%. "Teddy Bear Cut", "Breed Trim" and "Undercoat Reduction," are each requested in the range of around 10-7% of the time. Finally, the least frequently requested services are Nail Trim & Paint and Lion Cut, which are both requested less than 5% of the time. Overall, A pie chart is an effective tool for visually displaying the proportions of several categories in a dataset. In this case, the chart illustrates which pet grooming services are most popular and which are least popular, as well as which services are most requested by the owners.

FOR SCATTER PLOT

import pandas as pd

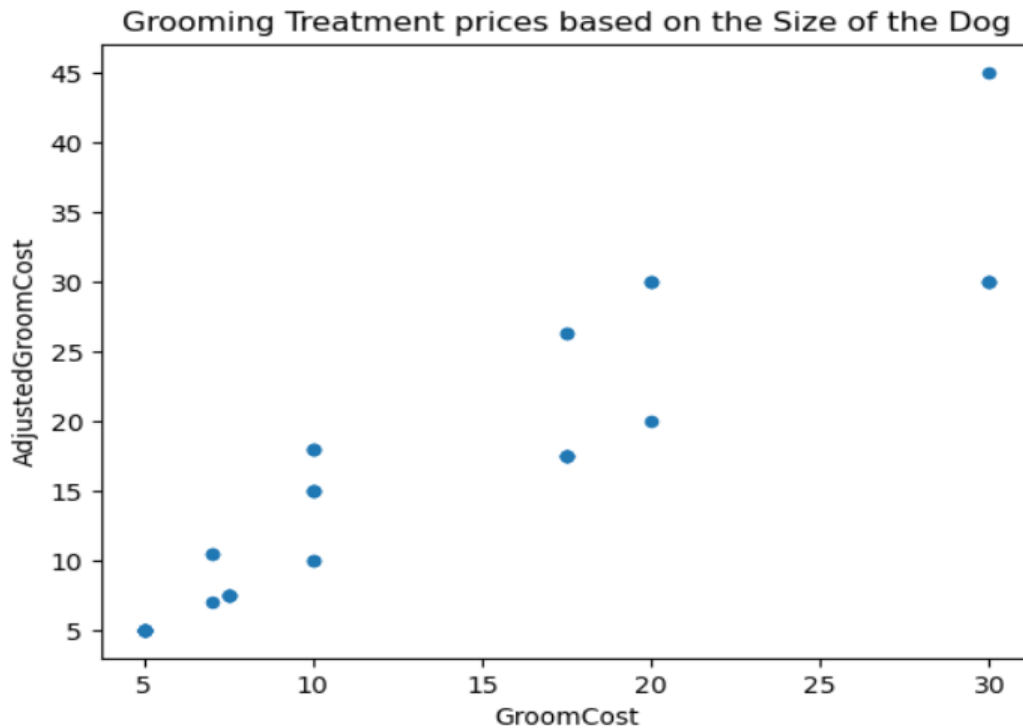
import pandas as pd

import matplotlib.pyplot as plt

data = pd.read_csv(r"C:\Users\shoai\Desktop\MutzNutzBookingSystem.csv")

data.plot.scatter(x = "GroomCost", y = "AdjustedGroomCost")

plt.title("Grooming Treatment prices based on the Size of the Dog")



A scatter plot is a type of graphical representation that uses points to represent the values of two different variables. Scatter plots are also commonly used to examine the relationship between two variables to determine if there is a correlation between them. Each point on the graph represents a pair of values, one for each variable. In this case, we have two variables: Groom Cost and Adjusted Groom Cost. The groom cost is the actual cost of grooming a dog, while the adjusted groom cost considers the size of the dog (small, medium, or large). The adjusted groom cost is to be higher for medium or larger dogs. However, the groom cost remains the same for some of the treatments. Additionally, because the "Adjusted Groom Cost" variable is based on dog size, we can use the scatter plot to see if there is a relationship between dog size and grooming costs. We can examine if larger dogs have higher grooming costs than smaller dogs, or whether there is no relationship between the two factors. Based on the graph, The x-axis would represent "Groom Cost," and the y-axis for "Adjusted Groom Cost." Each graph point represents a single observation (i.e., the groom cost and the corresponding adjusted groom cost). Overall, we can see that most of the data points fall along a diagonal line from the bottom left to the top right of the graph. This suggests a significant positive link between the two variables. As Groom Cost rises, so does Adjusted Groom Cost, and vice versa. Yet, there are a few data points that contradict this trend. A scatter plot

is a useful tool for graphically examining the connection between two variables and identifying any trends or outliers in the data.

Conclusion and Learning outcome

Discover about the specific strategies for constructing a relational database and learned how to construct SQL queries and how to visualize data.

REFERENCES

Hintea,D. D. (2023). *Normalization Week#3*.

https://files.coventry.aula.education/9033938a25326e5463e36a6e49365c02lecture_2c.mp4

[4](#)

Hintea,D. D. (2023). *SQL codes Week #4*.

[https://files.coventry.aula.education/e4ba1beb6b5ce553e38e3da7779bdc54sqlite_demo.m](https://files.coventry.aula.education/e4ba1beb6b5ce553e38e3da7779bdc54sqlite_demo.mp4)

[p4](#)

Hintea, D. D. (2023a). *SQL functions Week #5*.

[https://files.coventry.aula.education/2a58f6c256f6f45613b6d4083e94bb62sqlite_demo_2](https://files.coventry.aula.education/2a58f6c256f6f45613b6d4083e94bb62sqlite_demo_2.mp4)

[.mp4](#)

Hintea, D. D. (2023). *Demo on dataset statistics in Anaconda/Python*.

[https://files.coventry.aula.education/828a12b6e070b13f7f6dd5bf6f09a45bdemo_jupiter](https://files.coventry.aula.education/828a12b6e070b13f7f6dd5bf6f09a45bdemo_jupyter_notebook.mp4)

[notebook.mp4](#)

