



UNIVERSITI
TEKNOLOGI
PETRONAS

LAB WEEK 9

SEP 2025 SEMESTER

'AFFAN NAJIY BIN RUSDI

22010453

BACHELOR OF COMPUTER SCIENCE

DATA SCIENCE

TEB2164

Contents

Code.....	3
Output	5

Code

Activity 1

```
#Lab 7B Activity 1

library(readxl)
library(dplyr)
#library(xlsx)

setwd("C:/Users/AFFAN/Documents/GitHub/DS-academic/DS_Lab7-W9/Lab_B")
list.files()

#Import
uforeport <- read_excel("uforeport.xls")
titanic <- read_csv("titanic.csv")
View(uforeport)
View(titanic)

#Count NA value
sum(is.na(uforeport)) #18028 missing
sum(is.na(titanic)) #177 missing

#Get NA positions
which(is.na(uforeport))
which(is.na(titanic))

#Count NA each column
print(sapply(uforeport, function(x) sum(is.na(x))))
print(sapply(titanic, function(x) sum(is.na(x))))

#Managing Empty Cells
dim(uforeport)
uforeport_cleaned = na.omit(uforeport)
dim(uforeport_cleaned)

dim(titanic)
titanic_cleaned = na.omit(titanic)
dim(titanic_cleaned)

#Rename Variable Name
names(uforeport_cleaned) = gsub(" ", "_", colnames(uforeport_cleaned))

#View Cleaned
View(uforeport_cleaned)
View(titanic_cleaned)
```

Activity 2

#Lab 7B Activity 2

```
#Titanic Insights
titanic_subset <- select(titanic_cleaned, embarked, class, survived) #Select columns
cherbourg_data <- filter(titanic_subset, embarked == "C", class == "Third") #Filter
Cherbourg Third Class
survival_rate <- summarize(cherbourg_data, Survival_Rate = mean(survived) * 100)
#Summarize survival
total_cherbourg <- filter(titanic_subset, embarked == "C") #Filter Cherbourg
third_class_pct <- mutate(total_cherbourg, Is_Third = class == "Third") #Add column to
check Third Class
third_class_pct <- summarize(third_class_pct, Third_Pct = mean(Is_Third) * 100) #Calc
percentage

View(head(cherbourg_data, 10))
View(survival_rate)
View(head(total_cherbourg, 10))
View(third_class_pct)

# UFO Insights (simpler use of verbs)
ufo_subset <- select(uforeport_cleaned, State, Shape_Reported) #Select columns
top_states<- group_by(ufo_subset, State) #Group by State
top_states_sightings <- summarize(top_states, Count = n()) #Count sightings per state
top_states_moresightings <- filter(top_states_sightings, Count > 1) #Filter states with
more than 1 sighting
top_states_desc <- arrange(top_states_moresightings, desc(Count)) #Arrange in descending
order

View(head(top_states, 10))
View(head(top_states_sightings, 25))
View(head(top_states_moresightings, 20))
View(head(top_states_desc, 10))
```

Output

Activity 2 Titanic Dataset

	embarked	class	survived
1	C	Third	1
2	C	Third	0
3	C	Third	0
4	C	Third	0
5	C	Third	0
6	C	Third	0
7	C	Third	1
8	C	Third	0
9	C	Third	0
10	C	Third	1

Diagram 1.1 Passengers from Cherbourg who are Third class

After creating a subset of the cleaned titanic data which included the variables of embarked, class, and survived. I filtered the data to only show the passengers from Cherbourg and was the third class. This is the head for the first 1- passengers.

	Survival_Rate
1	43.90244

Diagram 1.2 Survival Rate of Cherbourg Passengers

Then I summarized the survival rate of the Cherbourg passengers by calculating the mean then multiplied by 100%. The rate is as shown.

	embarked	class	survived
1	C	First	1
2	C	Second	1
3	C	First	0
4	C	First	0
5	C	Third	1
6	C	Second	1
7	C	First	1
8	C	First	0
9	C	Third	0
10	C	Third	0

Diagram 1.3 All Passengers from Cherbourg

Then, I listed the total passengers from Cherbourg without class. Displayed is the head of the data with the first 10 rows.

	Third_Pct
1	31.53846

Diagram 1.4 Passengers who are Third class from Cherbourg

Then I mutated the dataset to create a new column called `Is_Third` to specify the passengers who joined the third class while eliminating others that are not third class from the total passengers from Cherbourg as shown by Diagram 1.3. Thus, the percentage of passengers that are third class are shown as above.

Activity 2 UFO Report Dataset

	State	Shape_Reported
1	SC	SPHERE
2	AK	OTHER
3	VA	FORMATION
4	CA	SPHERE
5	CA	CHEVRON
6	CA	DISK
7	IN	SPHERE
8	AL	CIRCLE
9	CA	DISK
10	ID	CIGAR

Diagram 2.1 Grouped by State

I have created a subset as well for the ufo report by the states and shape it was identified as and grouped it by state. This here displayed the head of the data with the first 10 rows.

	State	Count
1	AK	23
2	AL	18
3	AR	24
4	AZ	130
5	CA	389
6	CO	57
7	CT	22
8	DE	4
9	FL	109
10	GA	42
11	HI	15
12	IA	26
13	ID	9
14	IL	100
15	IN	44
16	KS	27
17	KY	32
18	LA	16
19	MA	41
20	MD	32
21	ME	26
22	MI	68
23	MN	36
24	MO	79
25	MS	17

Diagram 2.2 Count of sighting per state

Then, I initiated the counting of the sightings by state. This is the first 25 rows of the countries with the sightings of the UFO.

	State	Count
1	AK	23
2	AL	18
3	AR	24
4	AZ	130
5	CA	389
6	CO	57
7	CT	22
8	DE	4
9	FL	109
10	GA	42
11	HI	15
12	IA	26
13	ID	9
14	IL	100
15	IN	44
16	KS	27
17	KY	32
18	LA	16
19	MA	41
20	MD	32

Diagram 2.3 Count of more than one sighting per state

To eliminate the states with no sightings and I wanted the states with more than one sighting (2, 3, 4, ...). I displayed the first 20 rows of the data.

	State	Count
1	CA	389
2	WA	187
3	AZ	130
4	TX	120
5	NY	111
6	FL	109
7	IL	100
8	PA	84
9	OH	83
10	MO	79

Diagram 2.4 Count of more than one sighting per state in descending order

After I got the states with more than one sighting, I decided to arrange it in descending order with the top 10 states with the most UFO sightings. The most sightings is California, followed by Washington and the third place goes to Arizona.

