Assignment - A

Q1) Data A excel → Transform data. **BOX PLOT**

By selecting only daily hire. ☐✓

→ close & apply (load)

→ click on daily hire (Right side)

→ Modeling (heading)

→ New column

1st clm ← Day of the week = FORMAT ('Daily Hire' [Day], "DDD")

Again New column,

2nd clm ← Day of the week key = Weekday ('Daily hire' [Day])

Again click on New measure,

Measures

- Maximum hire = MAX ('Daily Hire' [Number of Bicycle Hires])
- Minimum hire = MIN ('Daily Hire' [Number of Bicycle Hires])
- Mean Hires = Average ('Daily Hire' [Number of Bicycle Hires])
- Standard Dev = STDEV.P ('Daily Hire' [Number of Bicycle Hires])
- Total Hires = Sum ('Daily Hire' [Number of Bicycle Hires])

Choose Table from Visualizations.

↓ select day ☐✓ and total hires ☐✓ for that table.

File → options & settings → options → Off time Intelligence & Background data ☐ X

Choose one more New table

→ drag and add all the 5 measures

→ max hire, mean hire, min hire, SD, Total Hires

... Get more visuals

add your mail ID

→ Search - Box

Box & Whisker.

Add.

Category - day of the week

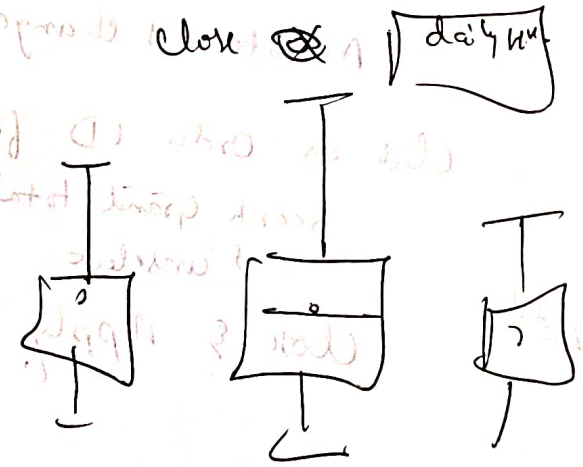
Sampling - Day

→ Model

Select day of the week

Advance.

Sort by dm → day of the week key



Ques 3) Data c ell:

Get data → data c → dirty 3 → Transform data

Remove × changed types × promot header.

Ctrl ~~Shift~~ select Colum 1 clm 2 → Right click → merge clms → comma → ok

click of Transform → Transpose

Use first row as header

Transform Tab → Fill → Down

Right click on 2 clms → unpivot other columns.

— Select Attribute clmn → Right click → Split clm → by delimeter

Attribute.1 → chang Name = Order ID

Attribute 2 → change Name = Order date

Click on Order ID filter ▽

Search Grand total.

→ unselect

Pending data

close & Apply.

Q4) Pareto → Data M. (table, clmn, measu)

or Get data → data M → sheet 1 ☐ → Transform data.

Duplicate colum → Item, Segmt, Sales (separately)

Item copy, Segmt copy → merge clms → custom → —

close & apply.

Chang sheet 1 nam into DATA

Modify → New table.

Summary = Summarize ['Data', Data[Blend], "Sum Sales", Sum(Data [Sales]))

Lendu the table,

Create Clm & Measure.

-) Item (1)
-) Segm (2)
-) Ranking (4)

-) Total Sales (3)
-) Cum total (5)
-) Cum Per. (6)

Select Table from Visualization,

Colu → drag blend, Sum Sales, Cum total, Rank, Cum %, Total Sales

... more opti, sort by → Sum of Sale.

Visuals → line & stacked clm ch

X axis → Item, Seg., blend.

Clm y axis = Sum of sum Sals

Line Y axis = Cum %

Visual

Data labels → On

Graph ... sort by - descending sum of sales.

Q5) Market Basket analysis → CSV file ~~8 mark~~ 8 mark

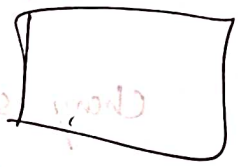
Get data → data f → Load.

Modelling → New table → Table = data f

Go to home → Text box → Type Market basket Analysis

From data f (Right side)

→ drag Item Name into Box.



Next, click on Stacked bar chart → from Visual.

Select Itetem Name from table □ (Right side)

X axis - Itemname

Y axis - Item name

Go to model view,

Connect bill No

Q94 Star Schema → data I → CSV

Get data → CSV → data I → Transform data.

Copy and Paste Orig data file → 3 times.

data(1) → Fact sales (change Name)
data(2) → dim_product
data(3) → dim customer
data(4) → dim_category

Except in fact sales, Remove duplicates from all files.

Got to fact sales :- choose clm → tab
→ Choose clm
Select only → Product ID, Custm ID, Units, unit cost, Unit price, date.

change ID headings into text.

After Index

Go to dim product → choose clm.
Product ID, Segmt, Categ., product

Merge Clmn dp → cat, seg
dcat → cat, seg.
unselect all OK
select segm seg → ok
Delete cat & seg clm

Click on Product ID, Right Click, Remove dupl.

Go to dim_cust → Custome ID, email, Zipcode, city, State, reg, district, Cowly.

ID & Zipcode → into text

Select Email name clmn → split Clmn → By delimiter → colon

Separate Name into First Name, Last name by split clmn

Email Name → Replace Value to remove ( ).

Go to dim_category → ~~Remove duplicates~~ Choose clm → category, segmt

Select both the columns and remove duplicates

Add clm → Index column ▾ From 1 change Index into Segmt Key ↔ Map them in dim_prod

Go to home → New source → blank Query ↓ Advance editory ↓ copy paste date text

Goto fact sales → date filter ▾
end ~~Start~~ date 30-06-2015
Start ~~end~~ date. - 30-07-2011

enter in query Start date & end date → Invoke.

Rename change that file → Invoke fun into Date.

Close & Apply

Go to Model

Connect date of date to fact sales date.

dim_cat category → dim-prod category.

Q7 → Hierarchy → data G.

Get data. → select all → Load

Go to model view

Click on Customer -
, * Country - Regi
Rightclick " Create hierach

Goto properties.
select → city, postal code, State - province,
customer

Change Name → Geography

Date → Create hierach → fiscal year, Quarter, date,
month

Product →. model, Category, Subcategory.

Reseller → County-Region → Create hier
City, postal code, State; provi, County-Reg,

Sales orders → sale orders, Sale order line

Region → Sales territories → ~~Region~~ count, group, Regn

Select Sales → New measure
↓
Sales amt by due date. = Calculate (Sum(Sales[Sales Amt]),

Date	Sales
date key	due date key
date key	Order date key
Date	Shift date key

Report view - Visualiz. -> Matrix

Product
- category -> Rows

Reseller
- Business type -> Rows

Sales ->
- sales amt -> values.

Visualiz -> cell elents

Data bone - on.

Select Map - Visualizati-

Reseller - country-region -> put in location

For Bubble size -> Sales -> Sum of order Qt

Select Staked area chart

Date -> fiscal ~~year~~ month -> x axis (Remove fiscal Qt, date)

Sales -> sum of sale amt -> Y axis (

Σ Sales Measure and amt

Q2) Power BI Intelligence Report - dataB

Select orders and data, transform.

→ Visualization

Card - Σprofit
Card - Σ Order Qty

Slicer - ~~ett~~ order date → data hier → Only Year

Q3 & Q8
Q5 & Q10 } Same

Q9 → Star Schem

→ take Screen
→ all 5 factsets
Entity relat → Excel
Star Sche → Power g
1 facts

AWS → hierarch

Python

pip install matplotlib
— Run —

```
from mpl_toolkits import mp
    ~~lot 3~~ lot3d
      import numpy as np
      import matplotib.pyplot as plt → Run
      import plotly.express as px
       df = px.data.iris()
        df.head(10) → Run (table)

        fig = px.scatter_3d(df,
                        x = 'sepal_length',
                        y = 'sepal_width',
                        z = 'petal_width',
                        color = 'species'
                        )
          fig.show() → Run
```

```
fig = px.histogram(df,
        x = 'sepal_length',
        nbins = 20,
        title = 'Distribution of Sepal lengths')
fig.show()
```