ASSIGNMENT 2 FRONT SHEET

Qualification	BTEC Level 5 HND Diploma in Computing		
Unit number and title	Unit 14: Business Intelligence		
Submission date		**Date Received 1st submission**	
Re-submission Date		**Date Received 2nd submission**	
Student Name	Nguyen Minh Son	**Student ID**	GCH18055
Class	GCH0715	**Assessor name**	Doan Trung Tung

Student declaration

I certify that the assignment submission is entirely my own work and I fully understand the consequences of plagiarism. I understand that making a false declaration is a form of malpractice.

Student's signature   Son

Grading grid

P3 P4 P5 P6 M3 M4 D3 D4

❑ **Summative Feedback:** ❑ **Resubmission Feedback:**

Grade: **Assessor Signature:**   **Date:**

IV Signature:

Table of Contents

Table of Figures :

Introduction

This report a detailed general concept about Business Intelligence, includes some tools / techniques for BI and their application in general. Also, I explain the dataset. Show how pre-process data for later analysis, explain each step and it purpose. Then design dashboards to show analysis on pre-processed data. Final is legal issues related to BI tools are presented in data analysis.

General concept

Definition: Business Intelligence is a set of processes, architectures, and technologies that convert raw data into meaningful information that drives profitable business actions.

BI has a direct impact on organization's strategic, tactical and operational business decisions. BI supports based on fact decision making using historical data rather than assumptions and feeling.

BI tools perform data analysis and create reports, summaries, dashboards, maps, graphs, and charts to provide users with detailed intelligence about the nature of the business.

Real examples of how to apply BI on business

Use BI to encourage customers to exercise and create good, long-term relationships with the company

- Filter customer data to classify types of customers based on their weight and Body Mass Index (BMI).

- Customers who reach their goal BMI and will be received gifts from the company.

- This will encourage customers to be motivated and create good relationships. Moreover, make customers feel good service with

the company. From there, customers will attach with the company for a long time

Use BI to filter top 10 best seller products.

- After filtering top 10 best seller products and top 10 products that have the lowest revenue, sales strategy will change.

- Company will focus on importing and selling products in the top 10 bestsellers. on the other hand, reduce import of products that sell less. In addition, products that have the lowest revenue, will boost sales by reducing product prices and using those products for events. Avoid products inventory.

Some techniques for BI

Analytics is a business intelligence technique that involves the study of available data to extract meaningful insights and trends.

Predictive modeling is a BI technique that utilizes statistical techniques to create models that could be used in forecasting probabilities and trends.

The model visualization technique is used to transform the discovered facts into histograms, plots, charts and other visuals that aid in proper interpretation of the insights.

Online analytical processing is a technique for solving analytical problems with different dimensions. The most important value in OLAP is its multidimensional aspect that lets users identify problems from different perspectives. OLAP could be used to complete tasks such as budgeting, CRM data analysis, and financial forecasting.

Some tools for BI

Sisens

If you're not tech-savvy, the Sisense BI tool could be a great option. It is incredibly user friendly and allows everyone within an organization to manage, analyze and visualize complex datasets without involving the IT department. This tool lets you gather data from various sources, including Google Analytics and AdWords. Since this tool uses in-chip technology, data processing is faster compared to other BI tools.

SAP Business Intelligence

SAP Business Intelligence provides an array of advanced analytics solutions including machine learning, BI predictive analytics, and planning and analysis. This enterprise-level applications for client/server systems offers data visualization and analytics applications, reporting and analysis, mobile analytics and office integration. SAP is a robust solution intended for all roles (management, end uses and IT) and offers a ton of functionalities in a single platform.

Dundas BI

Dundas BI is a browser-based BI tool that allows users to connect to multiple data sources in real-time. It provides great visualizations in tables, graphs and charts that could be customized and viewed from mobile devices and desktops. Users can easily build reports and extract certain performance metrics for purposes of analysis. Dundas offers support to all company types and across different industries.

(CallMiner, 2019)

Demonstration about BI

Datasets

This dataset has the survey data for the type of fitness practices that people follow.

List of the questions:

What is in the dataset

1. Name of the person attending the survey

2. Gender of the person attending the survey

3. Age of the person attending the survey

4. How important is an exercise to you on the scale of 1 to 5

5. How do you describe your current level of fitness? - Perfect, Very good, Good, Average, Un

6. How often do you exercise? - Every day, 1 to 2 times a week, 2 to 3 times a week, 3 to 4 ti

7. What barriers, if any, prevent you from exercising more regularly? (Select all that applies) - have an injury, I don't really enjoy exercising, I exercise regularly with no barriers

8. What forms of exercise do you currently participate in? (Select all that applies) - Walking o I don't really exercise

9. Do you exercise ? - Alone, With a friend, With a group, Within a class environment, I don't r

10. What time of the day do you prefer to exercise? - Early morning, afternoon, evening

11. How long do you spend exercising per day? - 30 min, 1 hour, 2 hours, 3 hours and above, I

12. Would you say, you eat a healthy balanced diet? - Yes, No, Not always

13. What prevents you from eating a healthy balanced diet, if any? (Select all that applies) - La have a balanced diet

14. How healthy do you consider yourself on a scale of 1 to 5?

15. Have you recommended your friends to follow a fitness routine? - Yes, No

16. Have you ever purchased fitness equipment? - Yes, No

17. What motivates you to exercise? (Select all that applies) - I want to be fit, I want to increas want to relieve stress, I want to achieve a sporting goal, I'm not really interested in exercisi

Figure 1: Datasets

Re-process data

Requires:

1. "timesamp" column is split into 2 columns:

- a column contains the date

- a column contains hours (does not contain GMT)

Timestamp
2019/07/03 11:48:07 PM GMT+5:30
2019/07/03 11:51:22 PM GMT+5:30

2. "Age" column is split into 2 columns

Example : 19 - 25 -> is split into 2 columns

Column 1: Min: 19

Column 2 : Max: 25

Note " Above " -> 100

Your age
19 to 25
19 to 25

3. "How often do you exercise? " column is split into 2 columns

Column 1 : How often do you exercise? (Min)

Column 2 : How often do you exercise? (Max)

Note: " ever day" = 7

"never" = 0

How often do you exercise?
Never
Never
1 to 2 times a week

4. In "How long do you spend exercising per day?" column, change hours into minutes (does not contain units).

Note : "I don't really exercise" = 0

How long do you spend exercising per day?
I don't really exercise
I don't really exercise
30 minutes

After cleaning:

1.

Date	Time
7/3/2019	11:48:07
7/3/2019	11:51:22

2.

Your age (min)	Your age (max)
19	25
19	25

3.

How often do you exercise? (Min)	How often do you exercise? (Max)
0	0
0	0
1	2

4.

How long do you spend exercising per day?
0
0
30

Code

First, we'll import the required libraries: CSV, Datetime and Regex (re):

```
1   #import libraries
2   import csv
3   import datetime as dt
4   import re
5
```

Open the original CSV file, we read it as key-value (dict reader):

```
6 ▼ with open('fitness-analysis.csv', mode='r', newline='') as csvfile: #open file
7       reader = csv.DictReader(csvfile) #parse CSV into key value pair
```

Create a new CSV file (result):

```
8
9       with open('result.csv', mode='w', newline='') as newfile: #write file
10
11
```

Create column (keys) for the new CSV file:

```
11          fieldnames = [
12          'Date',
13          'Time',
14          'Your name',
15          'Your gender',
16          'Your age (min)',
17          'Your age (max)',
18          'How important is exercise to you?',
19          'How do you describe your current level of fitness?',
20          'How often do you exercise? (Min)',
21          'How often do you exercise? (Max)',
22          'What barriers, if any, prevent you from exercising more regularly?',
23          'What form(s) of exercise do you currently participate in?',
24          'Do you exercise?',
25          'What time if the day do you prefer to exercise?',
26          'How long do you spend exercising per day?',
27          'Would you say you eat a healthy balanced diet?',
28          'What prevents you from eating a healthy balanced diet, If any?',
29          'How healthy do you consider yourself?',
30          'Have you ever recommended your friends to follow a fitness routine?',
31          'Have you ever purchased a fitness equipment?',
32          'What motivates you to exercise?'
33          ] #key value for new csv file
34          writer = csv.DictWriter(newfile, fieldnames=fieldnames) #write csv by key value
35          writer.writeheader() #write header (keys)
```

Then, read each row of the original CSV file and get the required values.

As for time format, we just separate the date and time:

```
37
38              for row in reader: |
39
40                  datetime = row['Timestamp'].split('GMT')[0].strip() #remove GMT
41                  datetimeFormat = dt.datetime.strptime(datetime, "%Y/%m/%d %H:%M:%S %p") #dat
42                  date = datetimeFormat.strftime('%Y/%m/%d')
43                  time = datetimeFormat.strftime('%H:%M:%S')
44
45
```

For age, we will split into two columns of lowest and highest age (min and max). We use Regex to get the age values, for example "19 to 25", the min value is 19, the max is 25. Or if the row has only min value (example: 40 and above) then the max value will be 100.

```
45
46                  age = row['Your age']
47                  ageMin = 0
48                  ageMax = 100
49                  reAge = re.search('^([0-9]+) to ([0-9]+)$', age) #using regex to search the
50                  if (reAge):
51                      result = re.findall('^([0-9]+) to ([0-9]+)$', age)[0] #returns a list co
52                      ageMin = result[0] #min is the first element in the list
53                      ageMax = result[-1] #max is the last element in the list
54                  else:
55                      result = re.findall('^([0-9]+) and above$', age)[0]
56                      ageMin = result
57
58          |
```

For the number of workouts per week, we also split into two columns min and max. Both the min and max columns will be 0 if the value in the row is "Nerver", and 7 if "Everyday". If there are exercise counts, continue to use Regex to get the min and max values.

```
58
59                  often = row['How often do you exercise?']
60                  oftenMin = 0 if often == 'Never' else 7 if often == 'Everyday' else re.finda
61                  oftenMax = 0 if often == 'Never' else 7 if often == 'Everyday' else re.finda
62
```

As for the time to exercise each time, it would initially be zero if he never went to exercise. Use Regex to check (search), if row has exercise time, switch to minutes format.

```
63
64                  exercisingTime = 0
65                  if (re.search('^([0-9]+) (.+)+$', row['How long do you spend exercising per
66                      result = re.findall('^([0-9]+) (.+)+$', row['How long do you spend exerc
67                      if (result[-1] == 'minutes'):
68                          exercisingTime = result[0]
69                      elif (result[-1] == 'hour' or result[-1] == 'hours'):
70                          exercisingTime = int(result[0]) * 60 #convert into minute
71
72          |
```

Finally, write the rows into the new CSV file as key-value:

```python
writer.writerow({
    'Date': date,
    'Time': time,
    'Your name': row['Your name'],
    'Your gender': row['Your gender'],
    'Your age (min)': ageMin,
    'Your age (max)': ageMax,
    'How important is exercise to you?': row['How important is exercise to you?'],
    'How do you describe your current level of fitness?': row['How do you describe y
    'How often do you exercise? (Min)': oftenMin,
    'How often do you exercise? (Max)': oftenMax,
    'What barriers, if any, prevent you from exercising more regularly?': row['What
    'What form(s) of exercise do you currently participate in?': row['What form(s) o
    'Do you exercise?': row['Do you exercise?'],
    'What time if the day do you prefer to exercise?': row['What time if the day do
    'How long do you spend exercising per day?': exercisingTime,
    'Would you say you eat a healthy balanced diet?': row['Would you say you eat a h
    'What prevents you from eating a healthy balanced diet, If any?': row['What prev
    'How healthy do you consider yourself?': row['How healthy do you consider yourse
    'Have you ever recommended your friends to follow a fitness routine?': row['Have
    'Have you ever purchased a fitness equipment?': row['Have you ever purchased a f
    'What motivates you to exercise?': row['What motivates you to exercise?']
}) #write rows to new csv file
```

Design Dashboard

Dashboard 1

Chart 1 :

This chart shows personal information for both men and women at all on the list. In the chart:

- Blue color represents the female gender
- Orange represents the male gender

Through this chart, we can identify the client's name, gender, and age, besides we can also identify how that person participates in sports activities. From there, it is possible to filter and identify those who intend to go to the gym
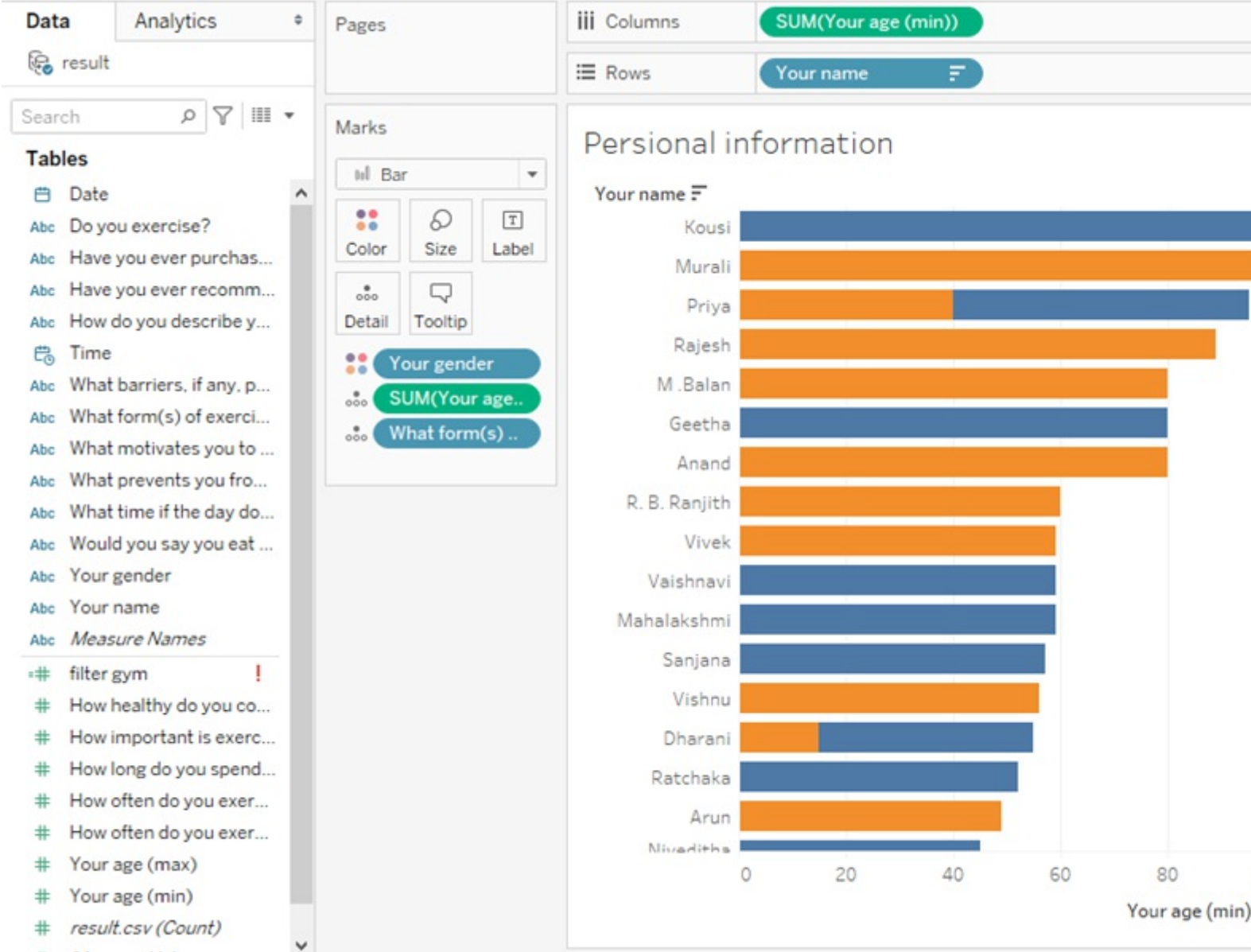
Figure 2: Customer personal information.

Chart 2: Health status of customers

This chart shows the health status of clients including male and female.

- Blue represents "No" answer to a healthy diet
- Orange represents the "Not always" answer to having a healthy diet
- Green represents the answer to "Yes" in having a healthy diet

Through this chart, we can determine the health status of the customer to see how much they are in addition to how they are physically able to show the information. news needed
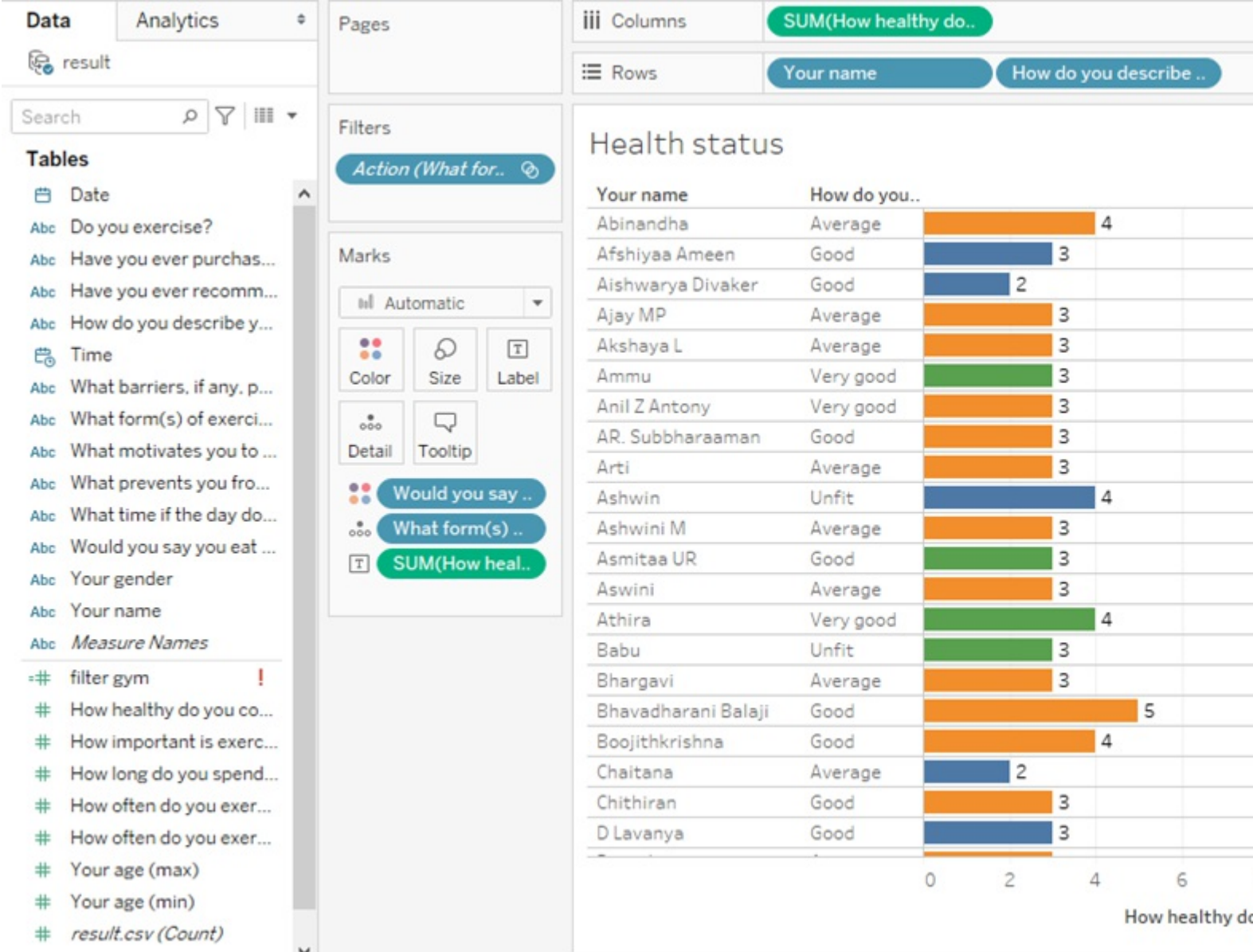
Figure 3: Health status of customers

Chart 3: Exercise habits of customers

In this chart, it will be more complicated because it will show 2 diagrams as "how often do you exercise?" with bar chart and "How long do you spend exercising per day?" with line chart style.

- The brown color represents those who have an afternoon workout routine
- Orange represents those who have an early morning routine
- Pink represents people who have a habit of practicing at night
- The blue line represents how long each household practices

Through this chart, one can look at their training history and how intense a client's exercise intensity will lead to many company-changing decisions.
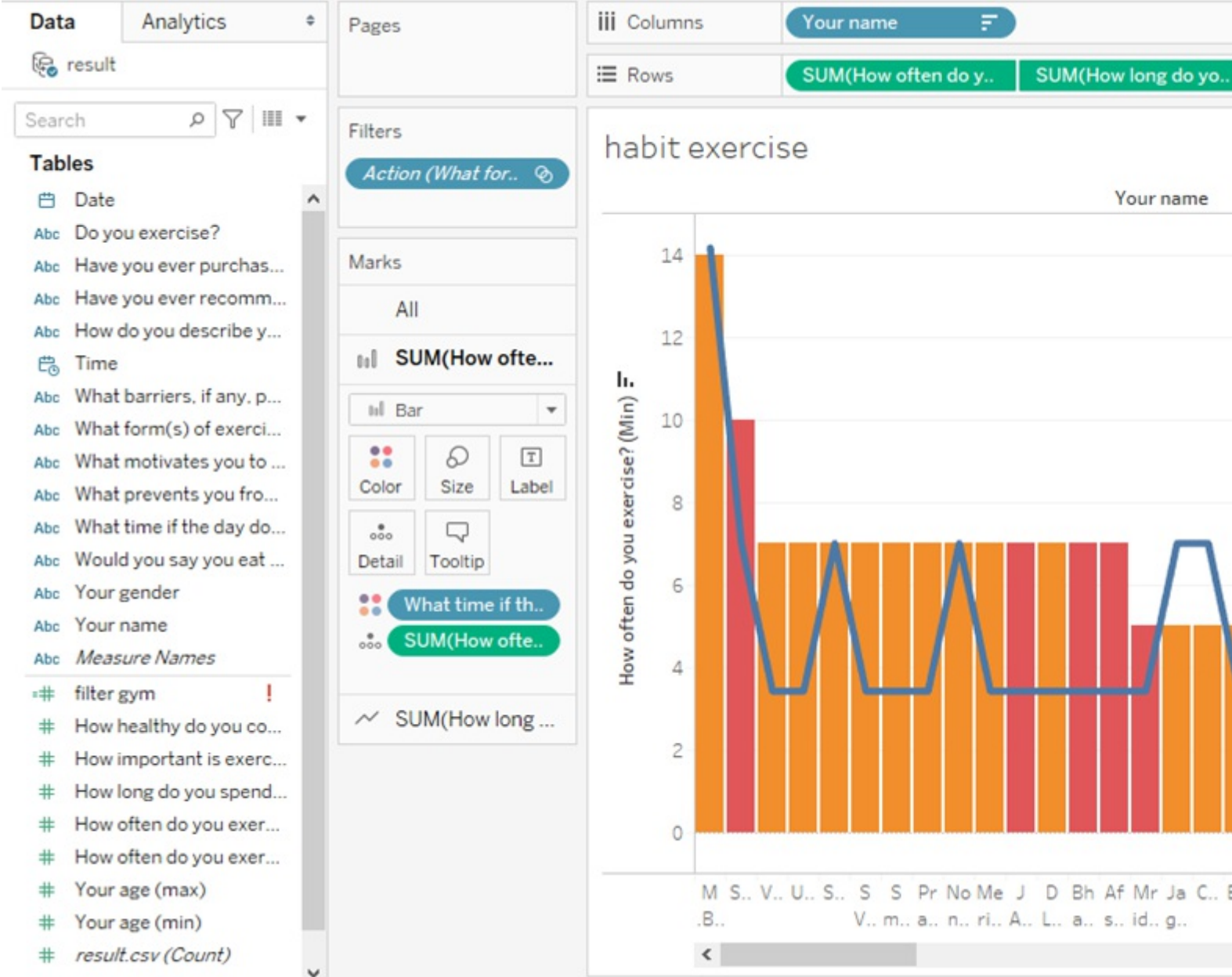
Figure 4: Exercise habits of customers

Dashboard

The dashboard aggregates everyone's data related to sports training and focus on those who work out. Besides, it will show information about each person, how they are in health and their training progress. From there, it is possible to come up with business strategies for the company and it motivates our company to take measures for those in need.

From this information, we can filter those with the same gym format to the same table to be able to track and provide analysis that affects the company.
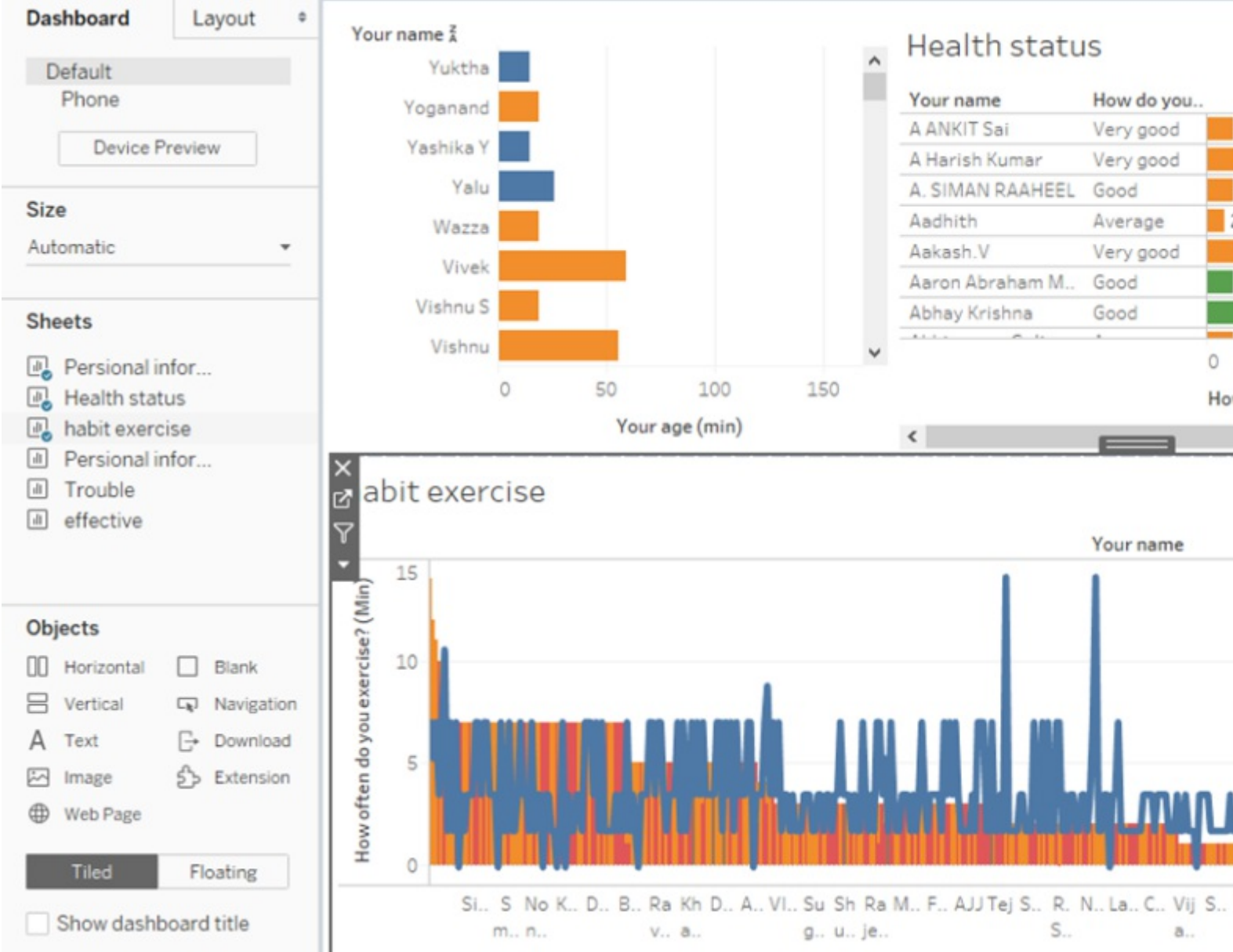
Figure 5: Dashboard 1

Dashboard 2

Chart 1 :

The personal information table (2) will store information of users including: name, max age, min age, gender and purpose of customers with the following attributes:

[Your name]

[Your gender]

[Your age (min)]

[Your age (max)]

[purpose]: filtered by code

if CONTAINS ([What motivates you to exercise?], 'I want to be flexible') then "I want to be flexible" ELSEIF CONTAINS ([What motivates you to exercise?], 'I want to relieve stress') then " I want to relieve stress "

ELSEIF CONTAINS ([What motivates you to exercise?], 'I want to lose weight') then "I want to lose weight"

ELSEIF CONTAINS ([What motivates you to exercise?], 'I want to achieve a sporting goal') then "I want to achieve a sporting goal"

ELSEIF CONTAINS ([What motivates you to exercise?], 'I want to be fit') then "I want to be fit"
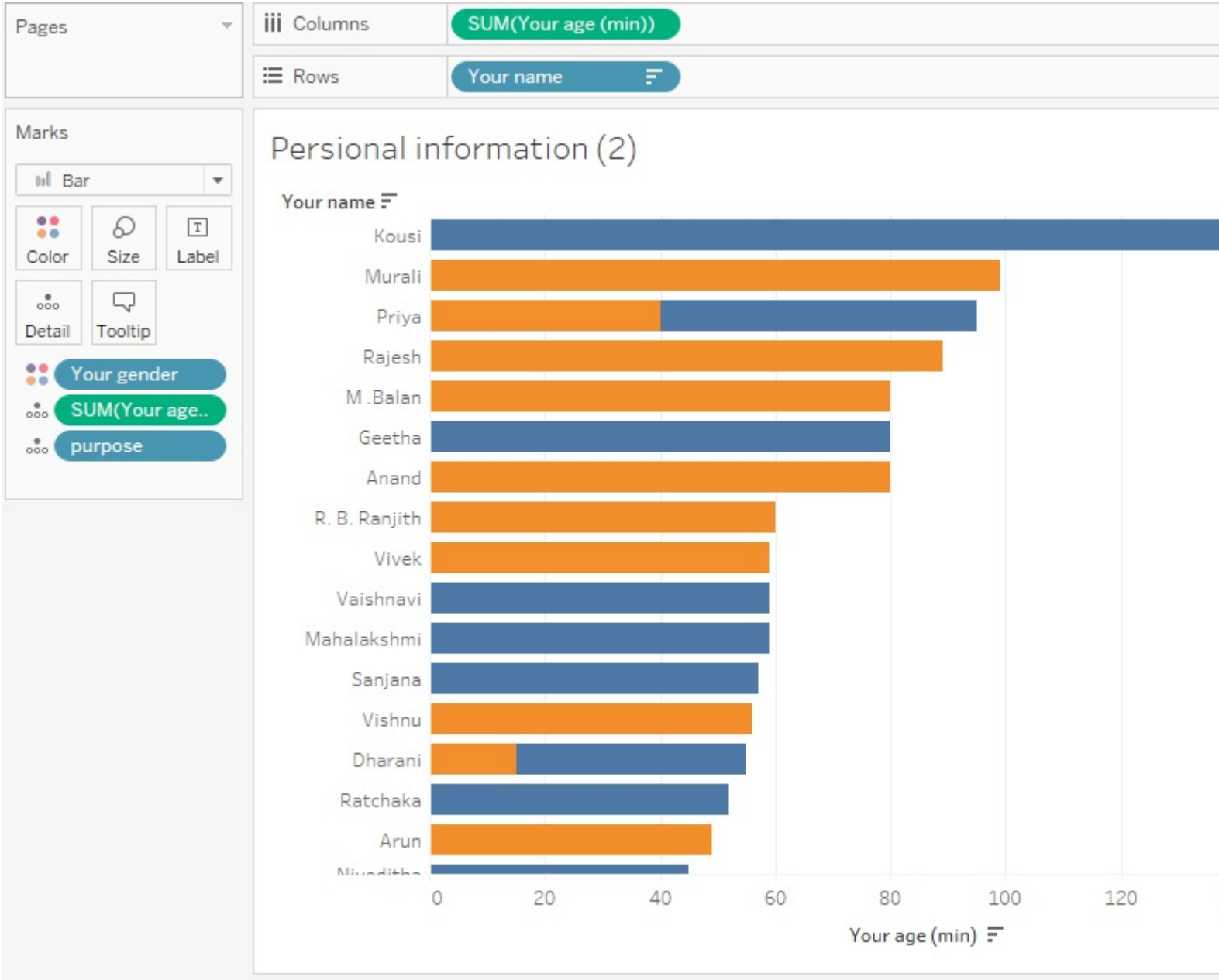
End

Figure 6:Personsal information

Chart 2 :

The Trouble table is a table that outlines a client's goals and the problems they encounter.

[Your name]

[Purpose]

[Trouble]: filtered by code

if CONTAINS ([What prevent you from eating a healthy balanced diet, If any?], 'Temptation and cravings') then "Temptation and cravings" ELSEIF CONTAINS ([What prevent you from eating a healthy balanced diet, If any?], 'Ease of access to fast food') then "Ease of access to fast food"

ELSEIF CONTAINS ([What prevent you from eating a healthy balanced diet, If any?], 'Cost') then "Cost"

ELSEIF CONTAINS ([What prevent you from eating a healthy balanced diet, If any?], 'Lack of time') then "Lack of time"

else "other"

end

[filter gym]: filtered by code

if CONTAINS ([What form (s) of exercise do you currently participate in?], 'Gym') then "Gym" else "Other"
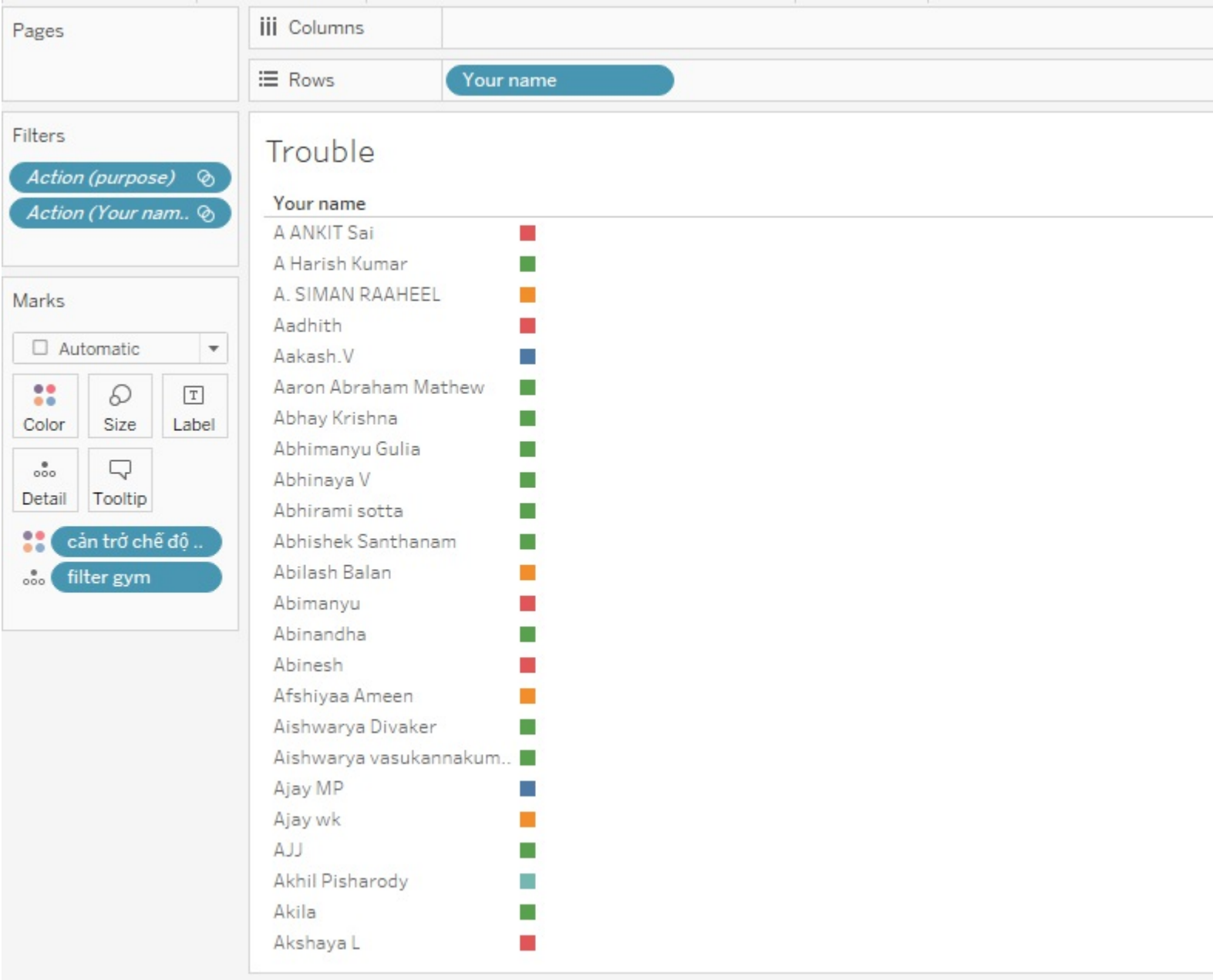
end

Figure 7: Trouble sheet

Chart 3 :

The effective table shows the results of practice of customers:

[Your name]

[How often do you exercise? (Min)]

[How often do you exercise? (Max)]

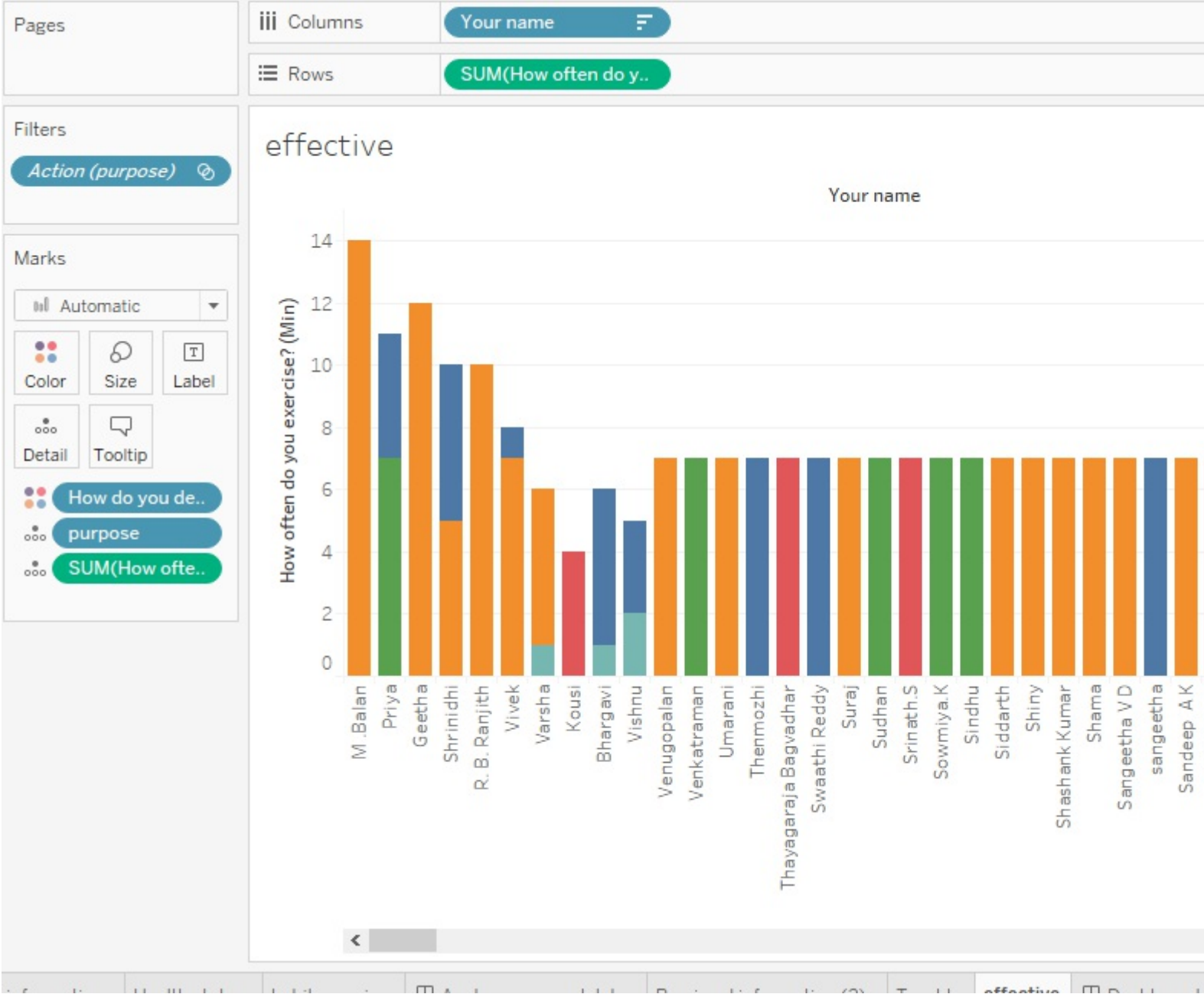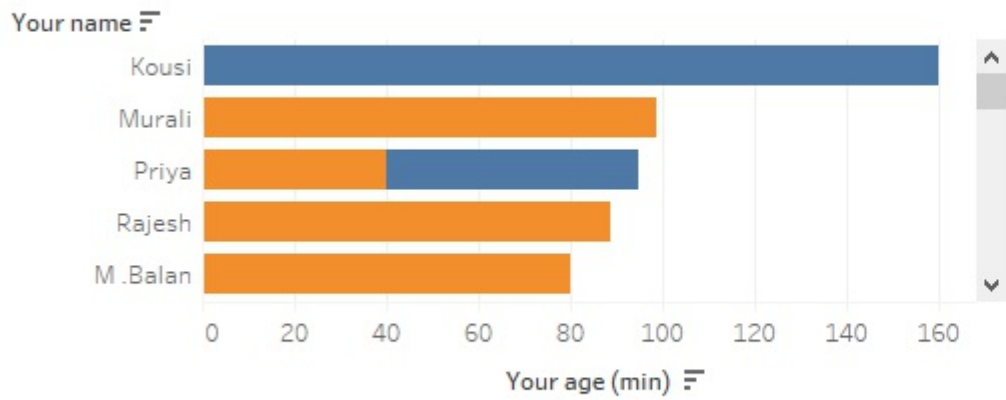[How do you describe your current level of fitness?]

[purpose]

Figure 8: Effective sheet

Dashboard:

Then, combine the 3 sheets above to create a Dashboard. Its meaning is to support clients who practice long and do not perform well. Then create a good relationship with customers.

Figure 9: Dashboard 1

Figure 10:Create relationship data

The Personal information sheet(2) is linked to the effective sheet via the [Purpose] field.

The effective sheet is linked to the Trouble sheet through the [Your name] field.

From the Personal information sheet(2), choose a training purpose. This will lead to changes to the effective table and the effective table will show practitioners for the chosen purpose: frequency of the practice and its effectiveness.

From there, find out those who practice a lot but are not effective. Selecting that person will change the Trouble panel and show all the obstacles that person is facing.

From there, the company can give people a 1-day free training to solve their problem. It both enhances the training atmosphere and creates good relationships with customers.

Review feedbacks

To take feedbacks and comments from viewers, I create Google Form to survey them.

Base on the data that was taken from Survey Online (Google form), I have analysis:
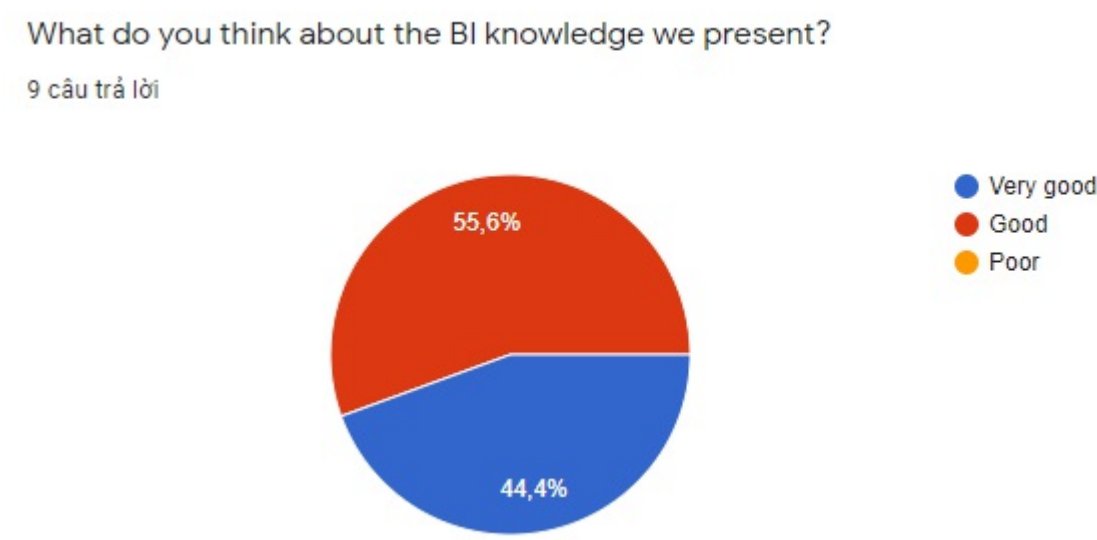

Figure 11: Question 1

55.6% of people think the BI knowledge we present is very good and 44.4% think it is good.
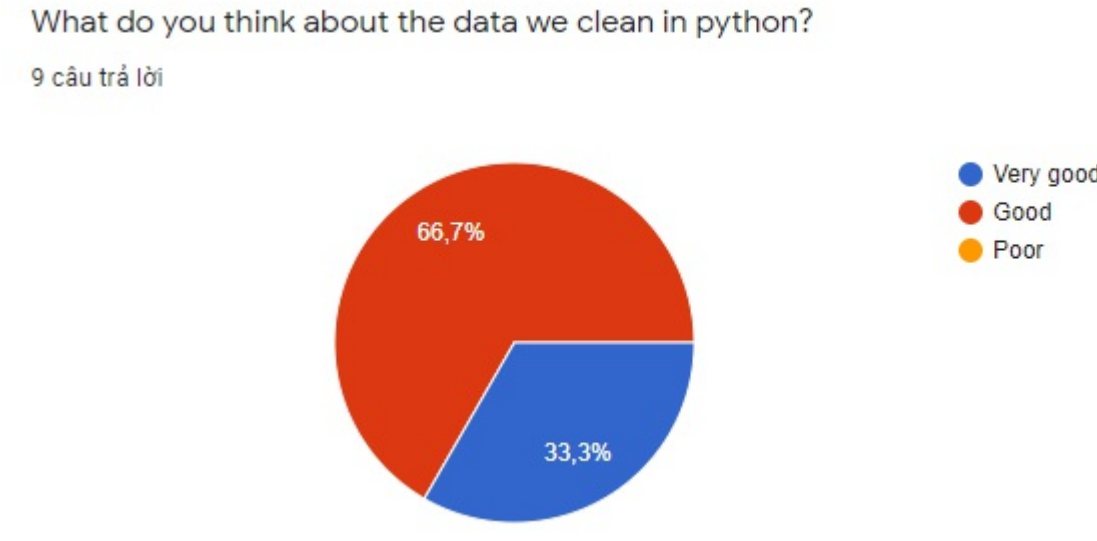

Figure 12: Question 2

33.3% of people think the data we clean up in python is very good and 66.7% think it is good.

How do you feel about our cleaning process?

9 câu trả lời

Legend: ● Very good ● Good ● Poor

44,4%

55,6%
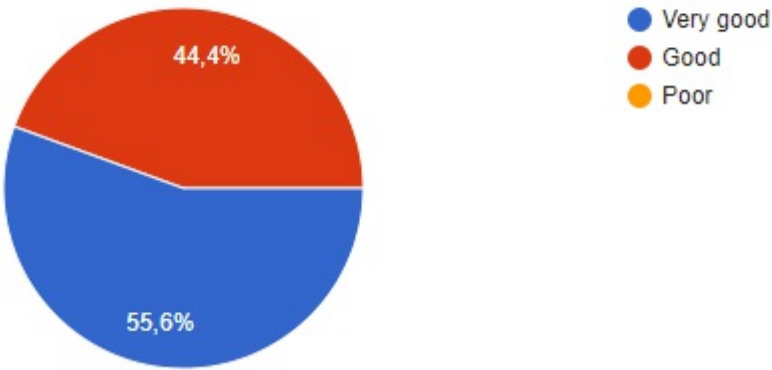
Figure 13: Question 3

55.6% of people think our cleaning process is very good and 44.4% think it is good.

How do you think about the graph that represents the data through the tableu?

9 câu trả lời

Legend: ● Very good ● Good ● Poor

55,6%

44,4%

Figure 14: Question 4

44.4% of people think that the graph that represents the data through the tableu is very good and 55.6% think it is good.

Evaluation of the group's presentation

9 câu trả lời

0 (0%) 0 (0%) 1 (11,1%) 6 (66,7%) 2 (22,2%)

1 2 3 4 5

Figure 15: Question 5

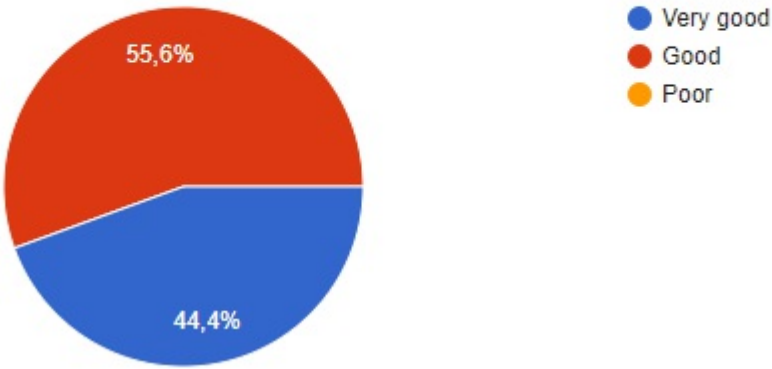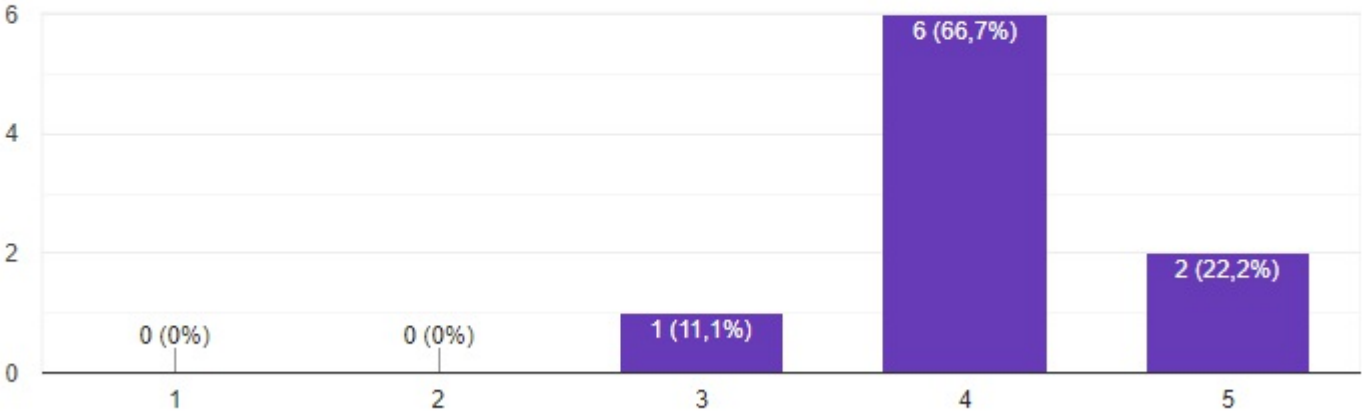11.1% rated the group presentation as 3/5, 66.7% rated the group presentation as 4/5, and 22.2% rated the group presentation as 5/5.

Through surveys we found we gave a successful presentation.

Feedback (if any)

5 câu trả lời

On dashboard 1, when searching for data on the gym, only find the gym and not show those who have gym, yoga, ..

no problem

No

You should analyze more carefully

it is OK

Figure 16: feedbacks

Problems:

On dashboard 1, when searching data about gym, only gym-only people are found, and not those who have exercised, yoga, etc.

There are some data that have not been applied yet, so they need to be included in the table for further analysis.

Solution problems:

We will use filter method in column (What form (s) of exercise do you currently participate in?) with the word gym in the answer all print out.

We will research more on this issue.

Improvements in the future:

- We will find more relationships between data to be able to open a new fitness related to increase revenue.

- We add questions about the customer's buying frequency so that we can find potential customers.

- We will create more charts of best-selling products to increase the revenue of best-selling products

The legal issues

Corporate legal departments have reclaimed from their law firms control of legal matters. As legal departments institute increasingly creative alternative fee arrangements to predict and control costs, they need objective data in order to negotiate rates. On the flip side of the coin, so too must law firms tap into actionable BI in order to assess factors such as trends in their historical billing rates, hours billed, etc. While legal departments may hold the reins in this relationship, firms must still engage in rational calculus to determine at which levels of service and compensation they can afford to be engaged, if at all.

Actionable BI leads to strategic decisions. Legal departments are better able to assess and manage risky matters. They can compare how effective individual attorneys are compared to colleagues in the same firm, as well as on an inter-firm basis. To the most effective go future spoils.

Legal departments have found that putting an e-billing and matter management system in place can save between five to 15

percent in legal fees solely by pre-screening invoices for accuracy and compliance with performance guidelines. As they learn to capitalize on the data-driven business intelligence solutions, legal departments have been able to drive savings of as much as 50 percent by making fully informed decisions data-driven outcomes.

(Kerschberg, 2011)

General Counsel or chief legal officers increasingly play a strategic role in advising board decisions, proactively managing information governance and compliance, and managing legal risk. As in-house legal teams assume greater responsibility for their dockets and remove that authority from their outside counsel, they need to find faster ways to improve data transparency to support decisions. Just as their own corporations implement innovative IT solutions to manage business more effectively, many chief legal officers are taking on the role of change agent, innovating with technology solutions in legal operations.

The flexible, high-performance legal operations they envision require a data-driven approach to decision-making, increasing demand for more complete solutions from law firms and other service providers. Corporate legal goals include greater use of technology tools, collection and analysis of management metrics, and project management training. (CHI, 2017)

Conclusion

The report demonstrated the use of business intelligence tools and technologies. Detail is using Tableau to demonstrate it. Moreover, get feedbacks from users to have improvements in the future.

References

CHI, S., 2017. *LawPay.* [Online]
Available at: https://www.lawpracticetoday.org/article/business-intelligence-better-legal-business-model/

Kerschberg, B., 2011. *Forbes.* [Online]
Available at: https://www.forbes.com/sites/benkerschberg/2011/04/26/business-intelligence-and-legal-matter-management/?sh=771cbec22d3c