BUSINESS INTELLIGENCE

LECTURE 1

OUTLINE

- Introduction
- Data mining overview
- Data pre-processing
- Classification techniques
- Clustering techniques
- Association rules
- References

INTRODUCTION

ALICE'S STORY

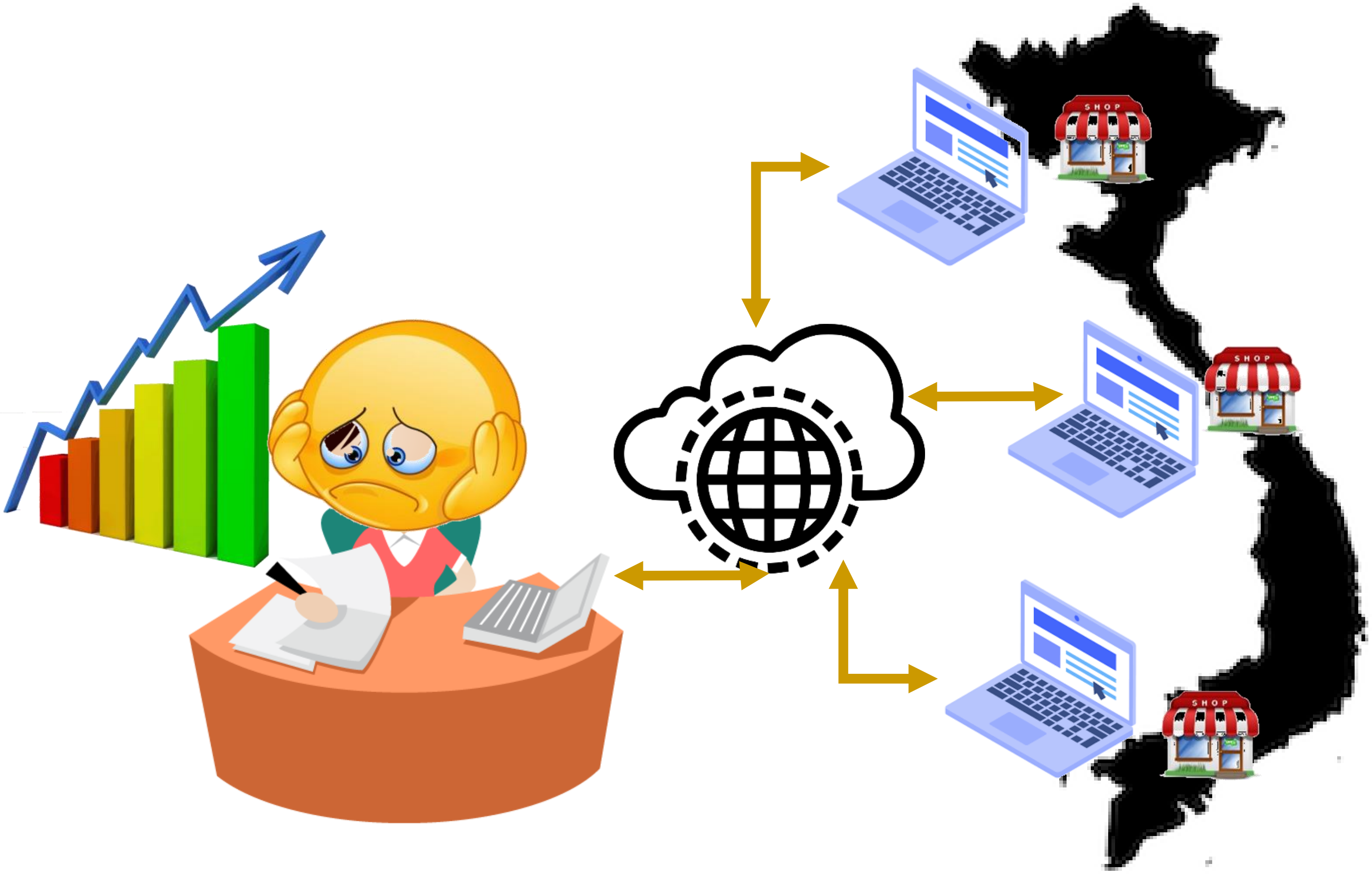
- Here is Alice
- She is interested in F&B
- One day, she opened her first store in Saigon
- It's a traditional restaurant with Vietnamese cuisine

http://www.dinh.vip/small-store-icon-psd-free-psd; https://iconbug.com/data/06/512/f91d30a37d1b631de0005b1402442924.png; https://static.pakwheels.com/2015/06/graph.gif

INTRODUCTION (1/5)

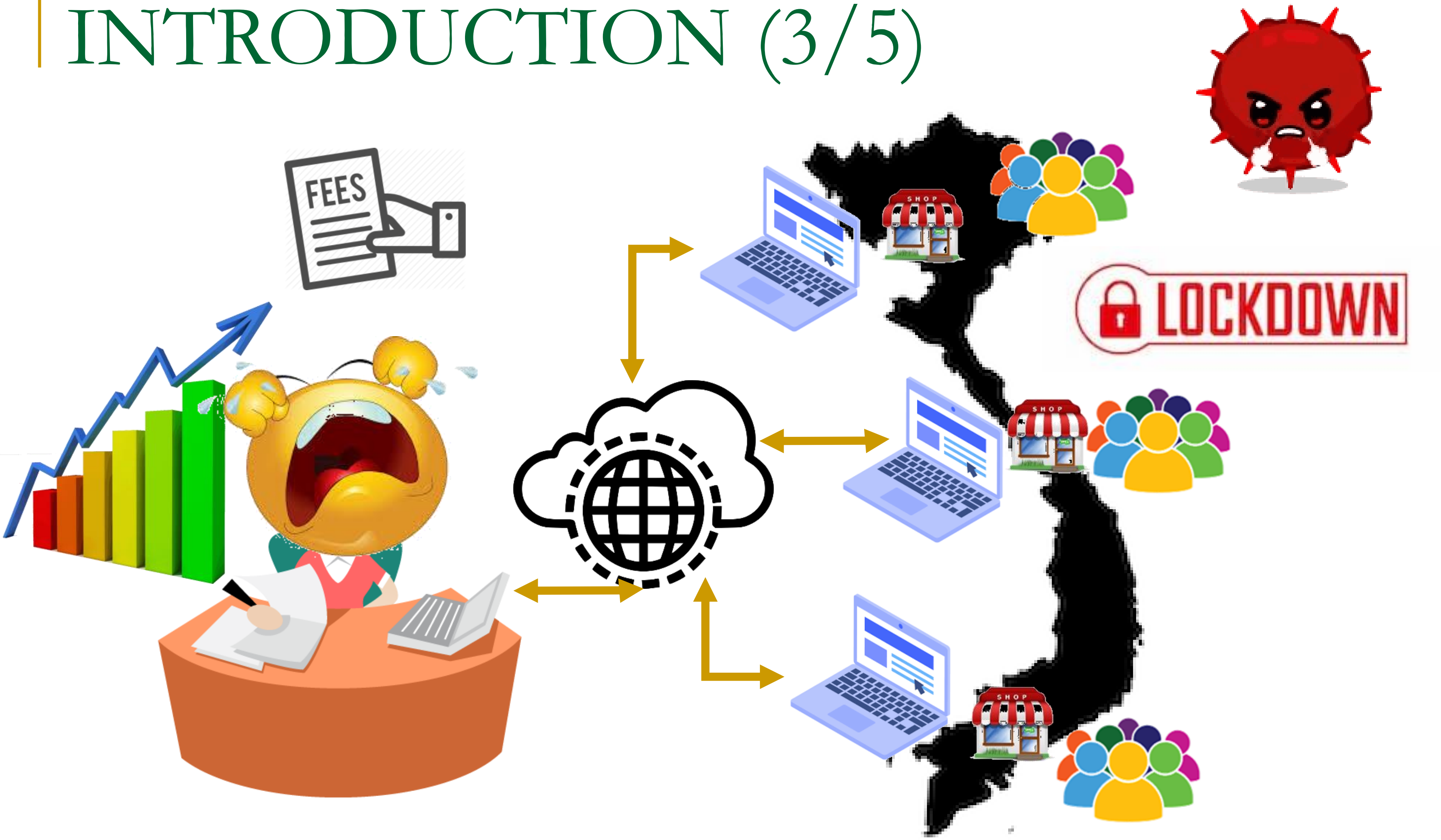
http://www.dinh.vip/small-store-icon-psd-free-psd; https://iconbug.com/data/06/512/f91d30a37d1b631de0005b1402442924.png; https://static.pakwheels.com/2015/06/graph.gif; https://cdn.pixabay.com/photo/2021/05/25/02/03/restaurant-6281067_1280.png;

INTRODUCTION (2/5)



http://www.dinh.vip/small-store-icon-psd-free-psd; https://iconbug.com/data/06/512/f91d30a37d1b631de0005b1402442924.png; https://static.pakwheels.com/2015/06/graph.gif; https://image.shutterstock.com/image-vector/high-detailed-vector-map-vietnam-260nw-150472901.jpg; https://encrypted-tbn0.gstatic.com/images?q=tbn%3AANd9GcQeeOZor21HxfRcdDaFqBdsNqBknZoYjo1b1h8Lv0FAmwXwLjqr&usqp=CAU; https://st2.depositphotos.com/1001911/7684/v/450/depositphotos_76840879-stock-illustration-depressed-emoticon.jpg; https://www.onlinewebfonts.com/icon/504308

INTRODUCTION (3/5)



http://www.dinh.vip/small-store-icon-psd-free-psd; https://iconbug.com/data/06/512/f91d30a37d1b631de0005b1402442924.png; https://static.pakwheels.com/2015/06/graph.gif; https://image.shutterstock.com/image-vector/high-detailed-vector-map-vietnam-260nw-150472901.jpg; https://encrypted-tbn0.gstatic.com/images?q=tbn%3AANd9GcQeeOZor21HxfRcdDaFqBdsNqBknZoYjo1b1h8Lv0FAmwXwLjqr&usqp=CAU; https://st2.depositphotos.com/1001911/7684/v/450/depositphotos_76840879-stock-illustration-depressed-emoticon.jpg; https://www.onlinewebfonts.com/icon/504308; https://encrypted-tbn0.gstatic.com/images?q=tbn%3AANd9GcRxN_w4kEKxFlrNl_P7sJrrIaKDbPgnxeleV3aZO7uCuRPPibJ8&usqp=CAU; https://telegramchannels.me/stickers/s-thevirus; https://www.iconfinder.com/icons/1955062/benefit_charge_fee_fees_finance_icon; https://www.shutterstock.com/search/lockdown+icon

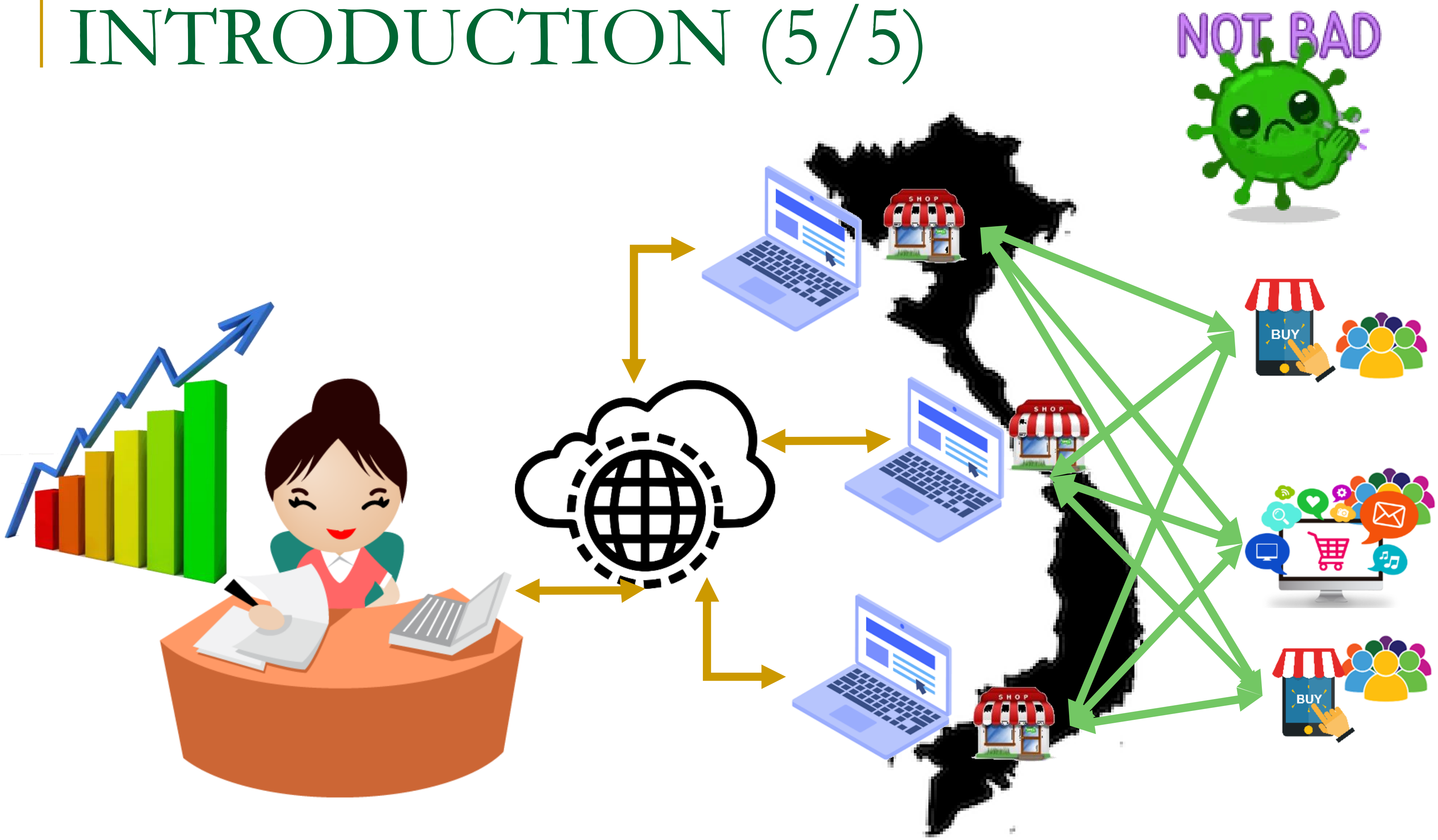
DISCUSSION

WOULD YOU SUGGEST ANY SOLUTIONS TO HELP HER?

INTRODUCTION (4/5)

https://violettek.com/wp-content/uploads/2018/09/violettek-e-store.png; https://www.paymill.com/en/blog/how-to-maintain-business-flow-through-m-commerce/

INTRODUCTION (5/5)

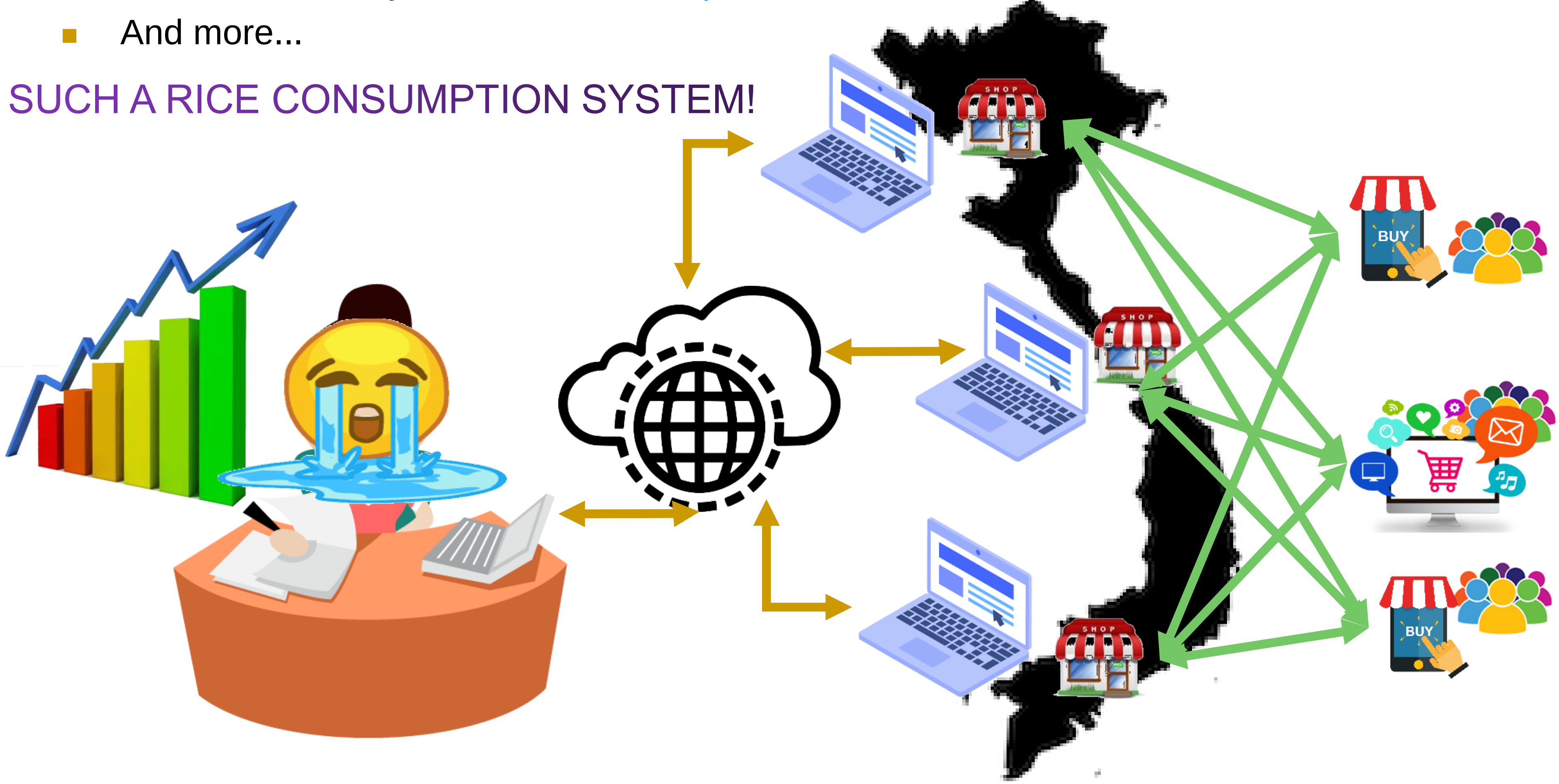


http://www.dinh.vip/small-store-icon-psd-free-psd; https://iconbug.com/data/06/512/f91d30a37d1b631de0005b1402442924.png; https://static.pakwheels.com/2015/06/graph.gif; https://image.shutterstock.com/image-vector/high-detailed-vector-map-vietnam-260nw-150472901.jpg; https://encrypted-tbn0.gstatic.com/images?q=tbn%3AANd9GcQeeOZor21HxfRcdDaFqBdsNqBknZoYjo1b1h8Lv0FAmwXwLjqr&usqp=CAU; https://st2.depositphotos.com/1001911/7684/v/450/depositphotos_76840879-stock-illustration-depressed-emoticon.jpg; https://www.onlinewebfonts.com/icon/504308; https://encrypted-tbn0.gstatic.com/images?q=tbn%3AANd9GcRxN_w4kEKxFlrNl_P7sJrrIaKDbPgnxeleV3aZO7uCuRPPibJ8&usqp=CAU; https://telegramchannels.me/stickers/s-thevirus

AND ANOTHER PROBLEM COMES…

- It takes her long time to summarize sales measurements
- She has to redo all for the new cycle (i.e., week, month, and year)
- When there is any error, she has to update her data and recheck the numbers
- And more...

SUCH A RICE CONSUMPTION SYSTEM!

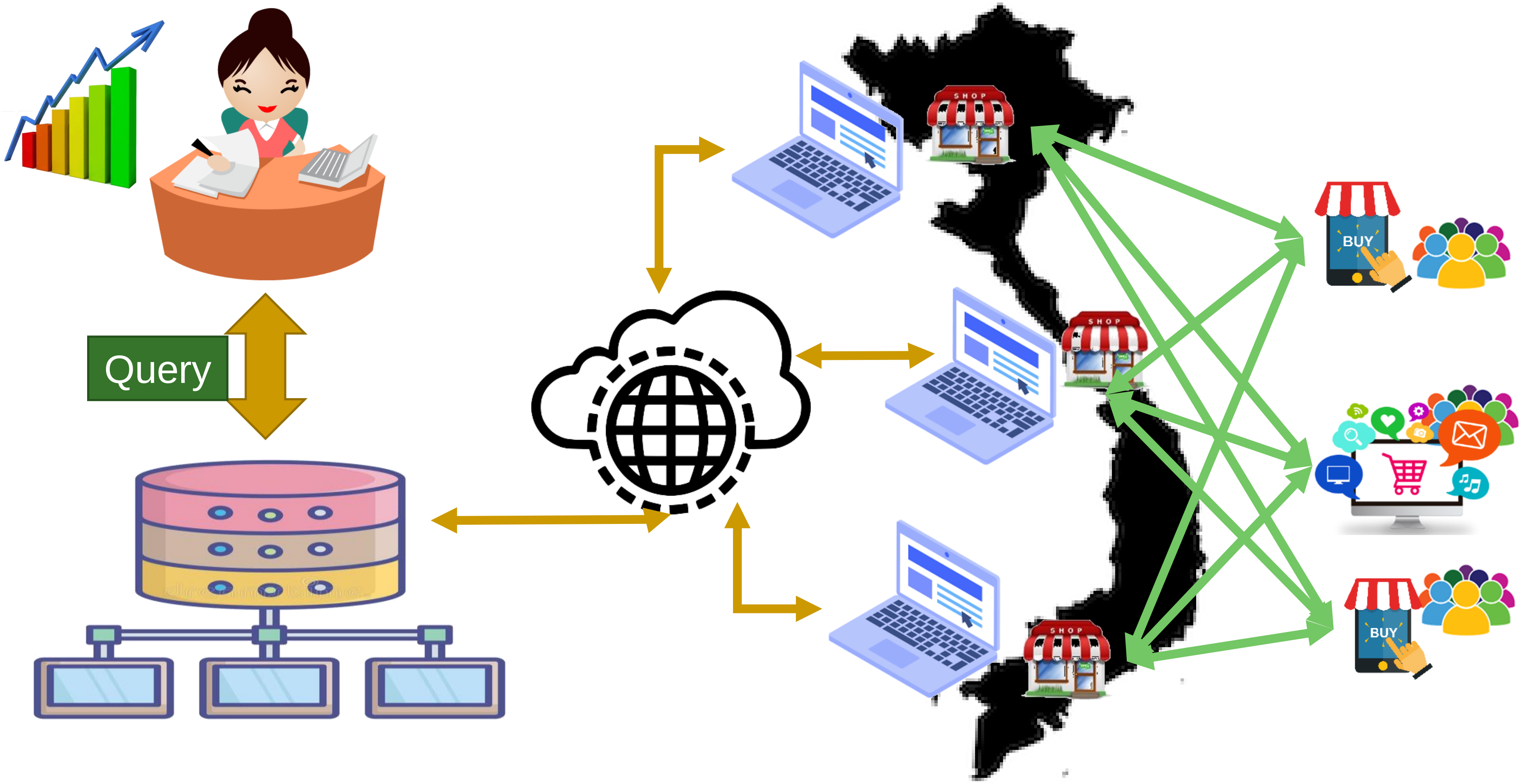


http://www.dinh.vip/small-store-icon-psd-free-psd; https://iconbug.com/data/06/512/f91d30a37d1b631de0005b1402442924.png; https://static.pakwheels.com/2015/06/graph.gif; https://image.shutterstock.com/image-vector/high-detailed-vector-map-vietnam-260nw-150472901.jpg; https://encrypted-tbn0.gstatic.com/images?q=tbn%3AANd9GcQeeOZor21HxfRcdDaFqBdsNqBknZoYjo1b1h8Lv0FAmwXwLjqr&usqp=CAU; https://st2.depositphotos.com/1001911/7684/v/450/depositphotos_76840879-stock-illustration-depressed-emoticon.jpg; https://www.onlinewebfonts.com/icon/504308; https://encrypted-tbn0.gstatic.com/images?q=tbn%3AANd9GcRxN_w4kEKxFlrNl_P7sJrrIaKDbPgnxeleV3aZO7uCuRPPibJ8&usqp=CAU; https://telegramchannels.me/stickers/s-thevirus; https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcQXP2CsFem1eg1ewxCqselmYS0lpmNJ7xRBAhtftDMzdfMBvZ0DfkGb7Zj6tjnIYEkQBuw&usqp=CAU

DISCUSSION

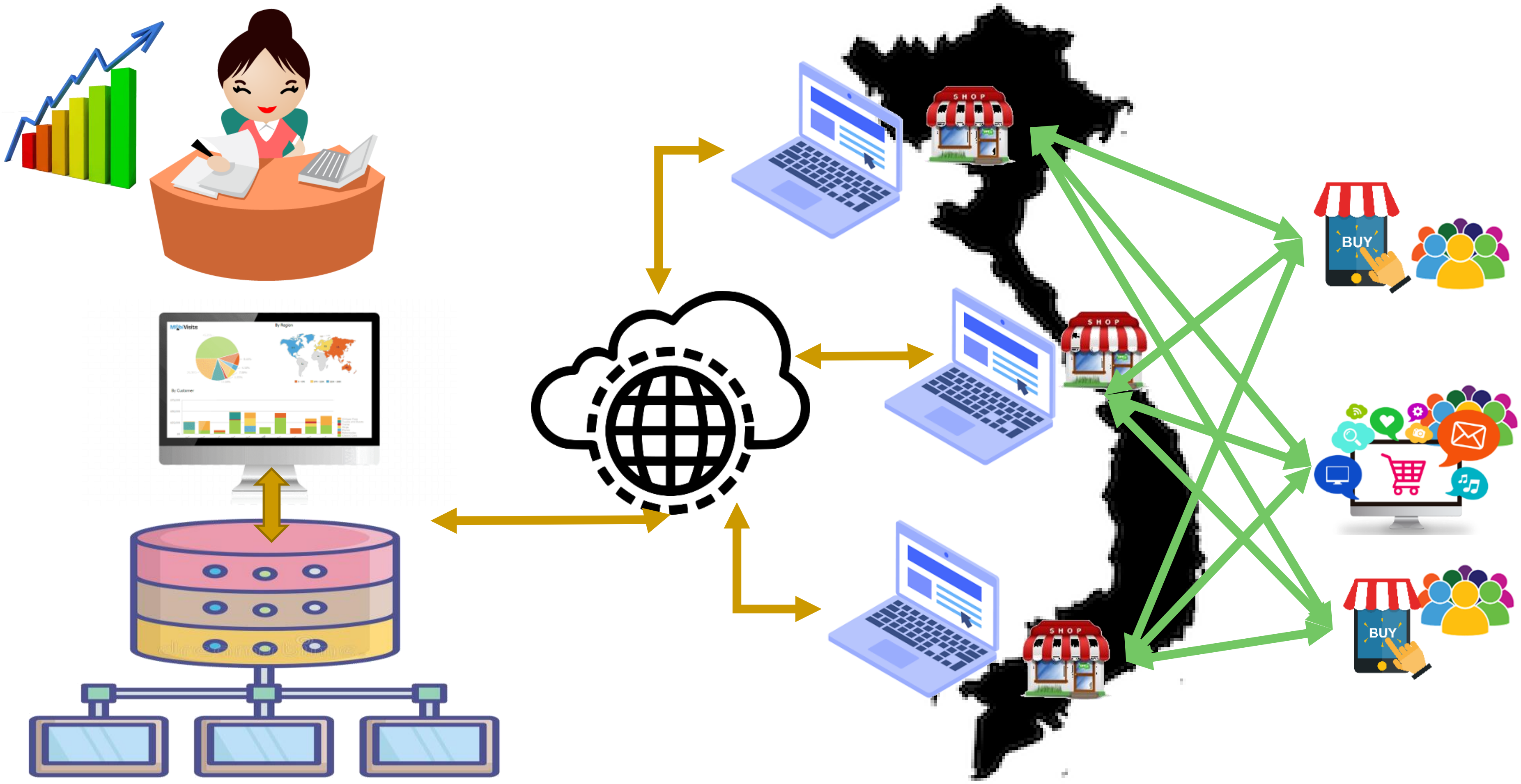
SHE NEEDS YOUR SUPPORT AGAIN!

YOU MAY COME UP WITH…



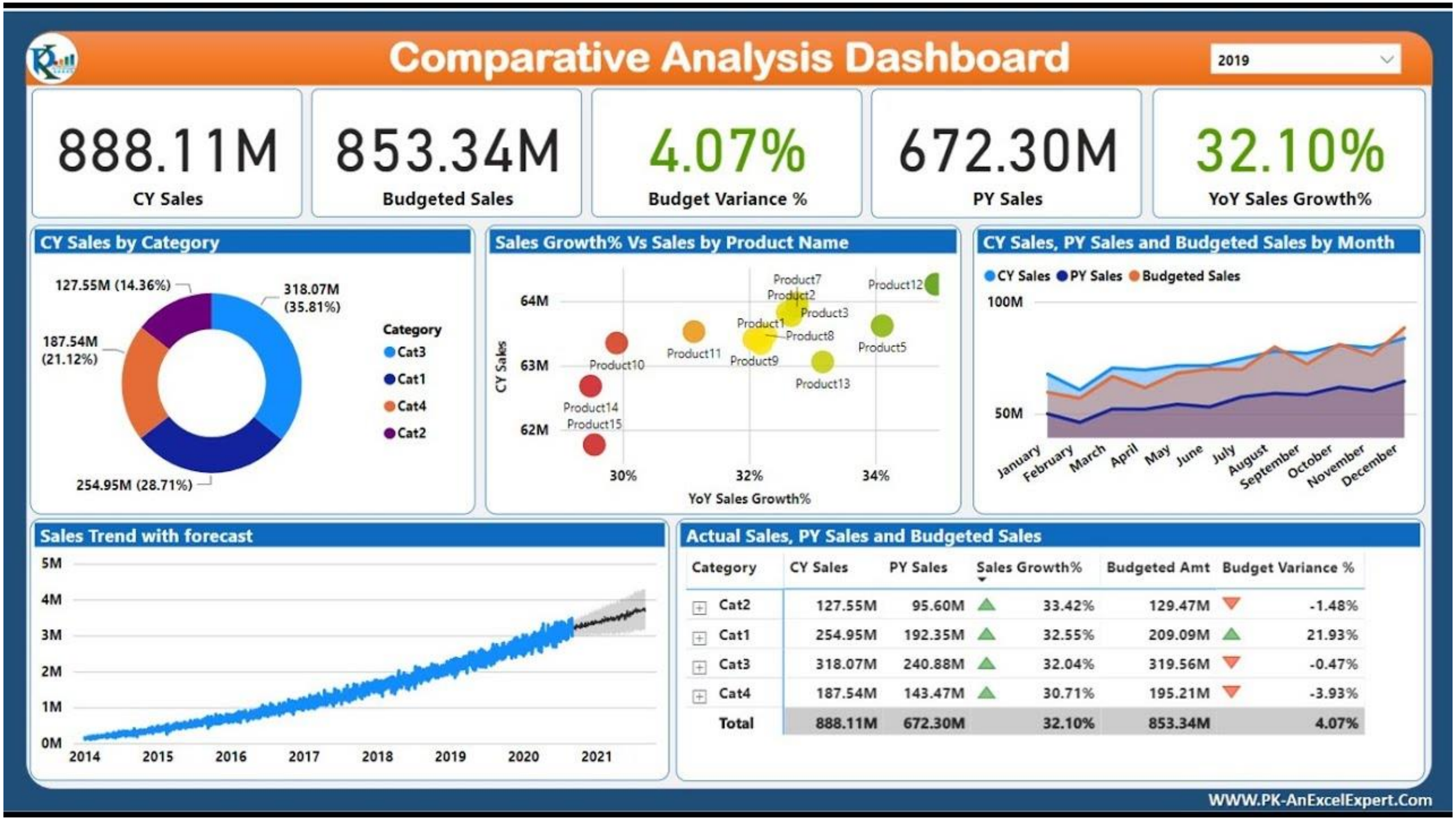
http://www.dinh.vip/small-store-icon-psd-free-psd; https://iconbug.com/data/06/512/f91d30a37d1b631de0005b1402442924.png; https://static.pakwheels.com/2015/06/graph.gif; https://image.shutterstock.com/image-vector/high-detailed-vector-map-vietnam-260nw-150472901.jpg; https://encrypted-tbn0.gstatic.com/images?q=tbn%3AANd9GcQeeOZor21HxfRcdDaFqBdsNqBknZoYjo1b1h8Lv0FAmwXwLjqr&usqp=CAU; https://st2.depositphotos.com/1001911/7684/v/450/depositphotos_76840879-stock-illustration-depressed-emoticon.jpg; https://www.onlinewebfonts.com/icon/504308; https://encrypted-tbn0.gstatic.com/images?q=tbn%3AANd9GcRxN_w4kEKxFlrNl_P7sJrrIaKDbPgnxeleV3aZO7uCuRPPibJ8&usqp=CAU; https://telegramchannels.me/stickers/s-thevirus; https://thumbs.dreamstime.com/b/centralized-database-fill-vector-icon-which-can-easily-modify-edit-centralized-database-fill-vector-icon-which-can-easily-185306424.jpg

OR BETTER FOR ALICE...



http://www.dinh.vip/small-store-icon-psd-free-psd; https://iconbug.com/data/06/512/f91d30a37d1b631de0005b1402442924.png; https://static.pakwheels.com/2015/06/graph.gif; https://image.shutterstock.com/image-vector/high-detailed-vector-map-vietnam-260nw-150472901.jpg; https://encrypted-tbn0.gstatic.com/images?q=tbn%3AANd9GcQeeOZor21HxfRcdDaFqBdsNqBknZoYjo1b1h8Lv0FAmwXwLjqr&usqp=CAU; https://st2.depositphotos.com/1001911/7684/v/450/depositphotos_76840879-stock-illustration-depressed-emoticon.jpg; https://www.onlinewebfonts.com/icon/504308; https://encrypted-tbn0.gstatic.com/images?q=tbn%3AANd9GcRxN_w4kEKxFlrNl_P7sJrrIaKDbPgnxeleV3aZO7uCuRPPibJ8&usqp=CAU; https://telegramchannels.me/stickers/s-thevirus; https://thumbs.dreamstime.com/b/centralized-database-fill-vector-icon-which-can-easily-modify-edit-centralized-database-fill-vector-icon-which-can-easily-185306424.jpg; https://e7.pngegg.com/pngimages/56/552/png-clipart-dashboard-business-intelligence-report-analytics-computer-icons-analysis-company-display-advertising.png

DASHBOARD



Comparative Analysis Dashboard

2019

888.11M
CY Sales

853.34M
Budgeted Sales

4.07%
Budget Variance %

672.30M
PY Sales

32.10%
YoY Sales Growth%

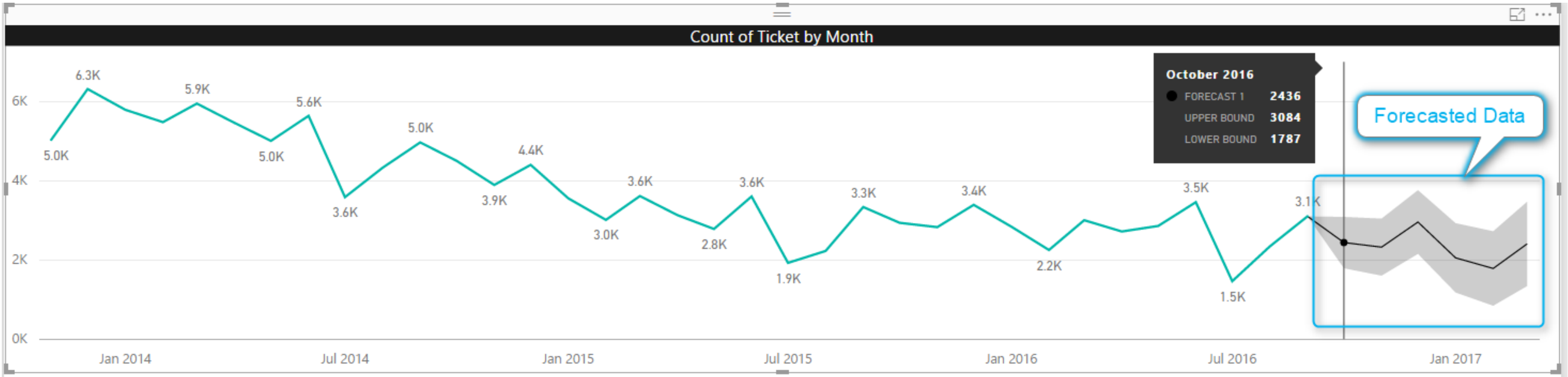
Actual Sales, PY Sales and Budgeted Sales

Category	CY Sales	PY Sales		Sales Growth%	Budgeted Amt		Budget Variance %
⊞ Cat2	127.55M	95.60M	▲	33.42%	129.47M	▼	-1.48%
⊞ Cat1	254.95M	192.35M	▲	32.55%	209.09M	▲	21.93%
⊞ Cat3	318.07M	240.88M	▲	32.04%	319.56M	▼	-0.47%
⊞ Cat4	187.54M	143.47M	▲	30.71%	195.21M	▼	-3.93%
Total	888.11M	672.30M		32.10%	853.34M		4.07%

WWW.PK-AnExcelExpert.Com

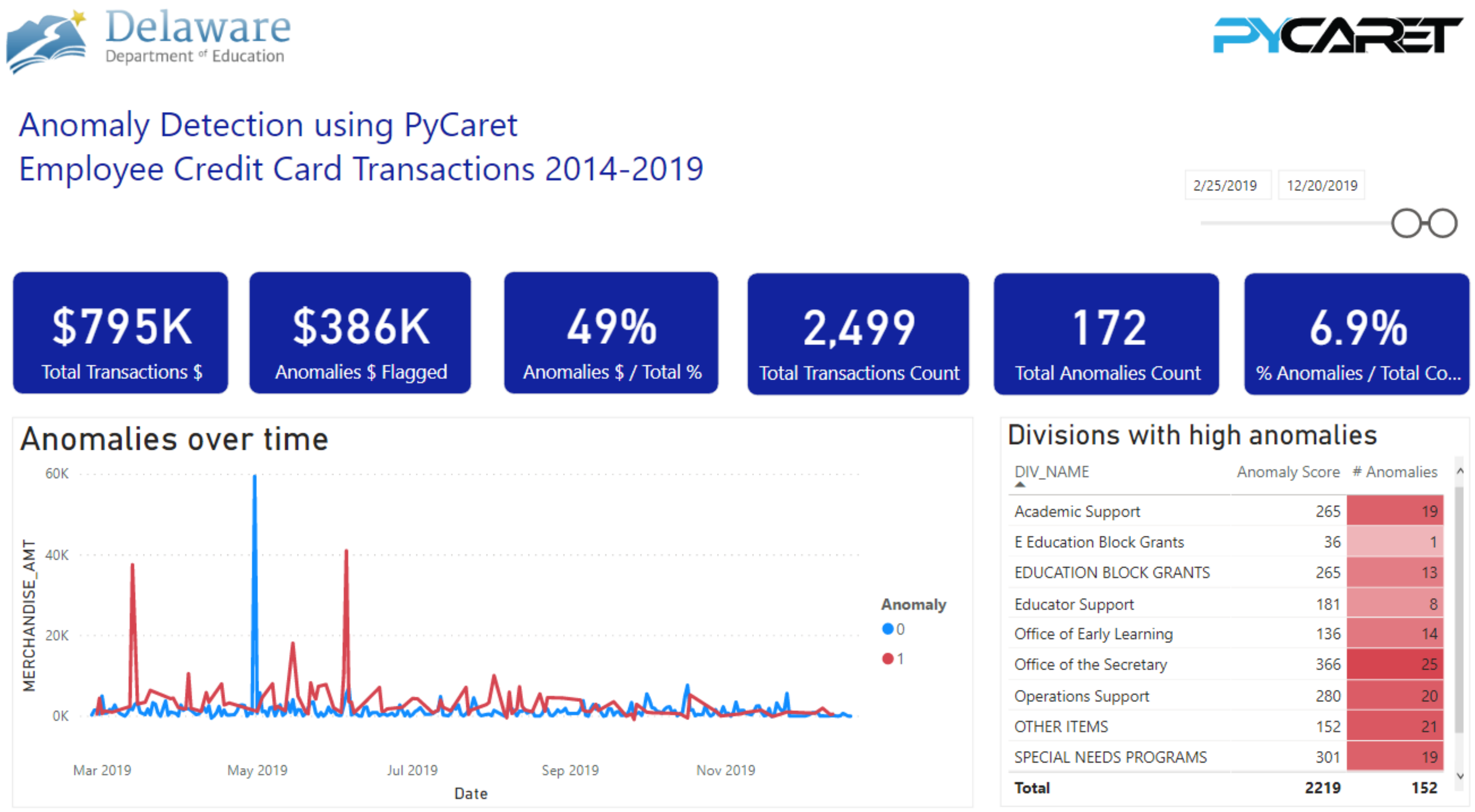
https://i.ytimg.com/vi/X7DsnK5bD-0/maxresdefault.jpg;

SALES FORECAST



https://databear.com/forecasting-in-power-bi/

ANOMALY DETECTION



Delaware Department of Education

PYCARET

Anomaly Detection using PyCaret
Employee Credit Card Transactions 2014-2019

2/25/2019 12/20/2019

$795K	$386K	49%	2,499	172	6.9%
Total Transactions $	Anomalies $ Flagged	Anomalies $ / Total %	Total Transactions Count	Total Anomalies Count	% Anomalies / Total Co...

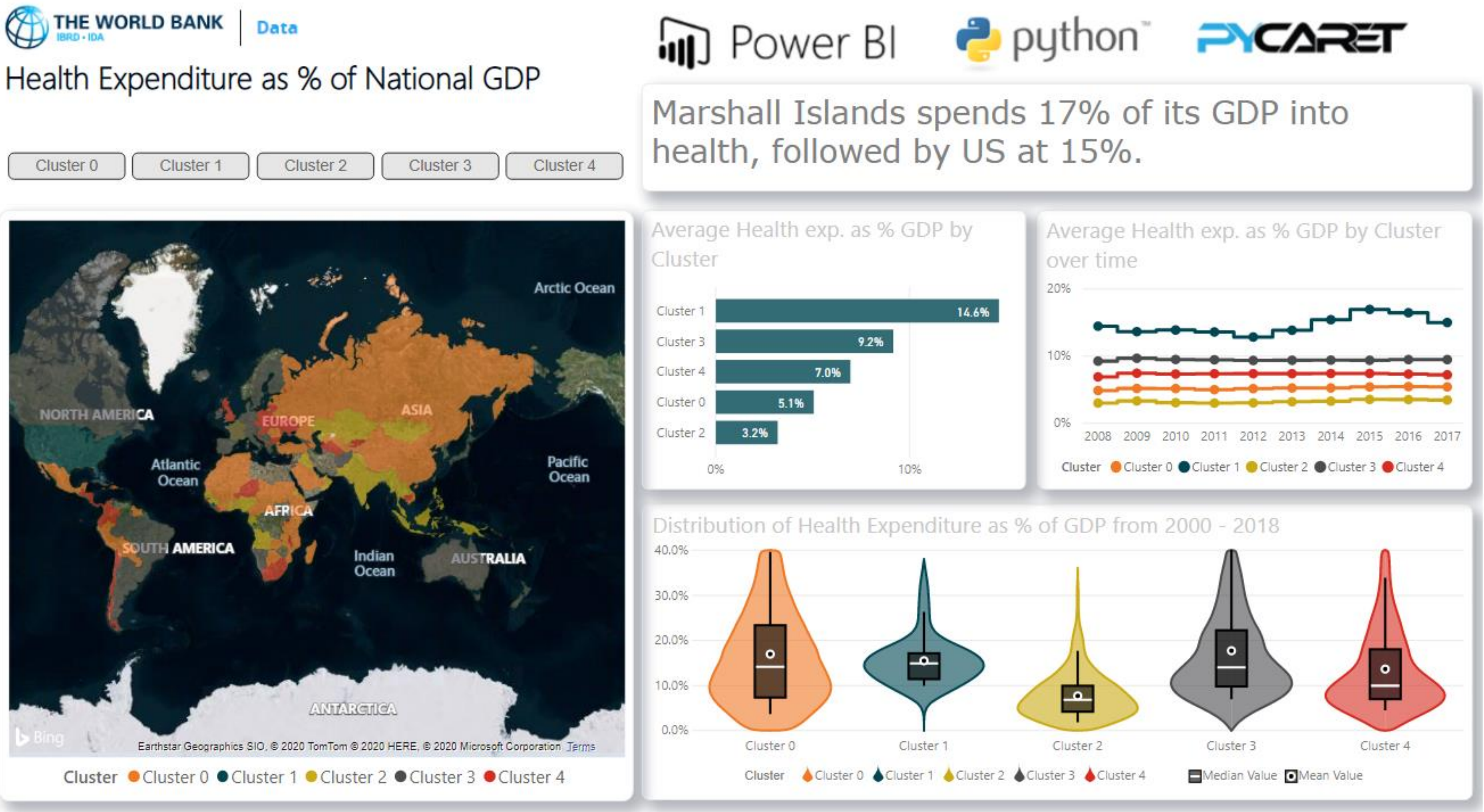
Anomalies over time

Divisions with high anomalies

DIV_NAME	Anomaly Score	# Anomalies
Academic Support	265	19
E Education Block Grants	36	1
EDUCATION BLOCK GRANTS	265	13
Educator Support	181	8
Office of Early Learning	136	14
Office of the Secretary	366	25
Operations Support	280	20
OTHER ITEMS	152	21
SPECIAL NEEDS PROGRAMS	301	19
Total	**2219**	**152**

https://miro.medium.com/max/1400/1*7qWjee_M6PTrAd0PJdU1yg.png

CLUSTER IDENTIFICATION



Health Expenditure as % of National GDP

Marshall Islands spends 17% of its GDP into health, followed by US at 15%.

https://miro.medium.com/max/1400/1*sUeqYcENVII1RlyYA_-Uxg.png

BI SYSTEM

- “Business intelligence is an infrastructure that helps in the process of collecting, storing, and analyzing data from business operations.
- BI provides comprehensive business metrics, in near-real-time, to support better decision making. You can create performance benchmarks, spot market trends, increase compliance, and improve almost every aspect of your business with better business intelligence.”

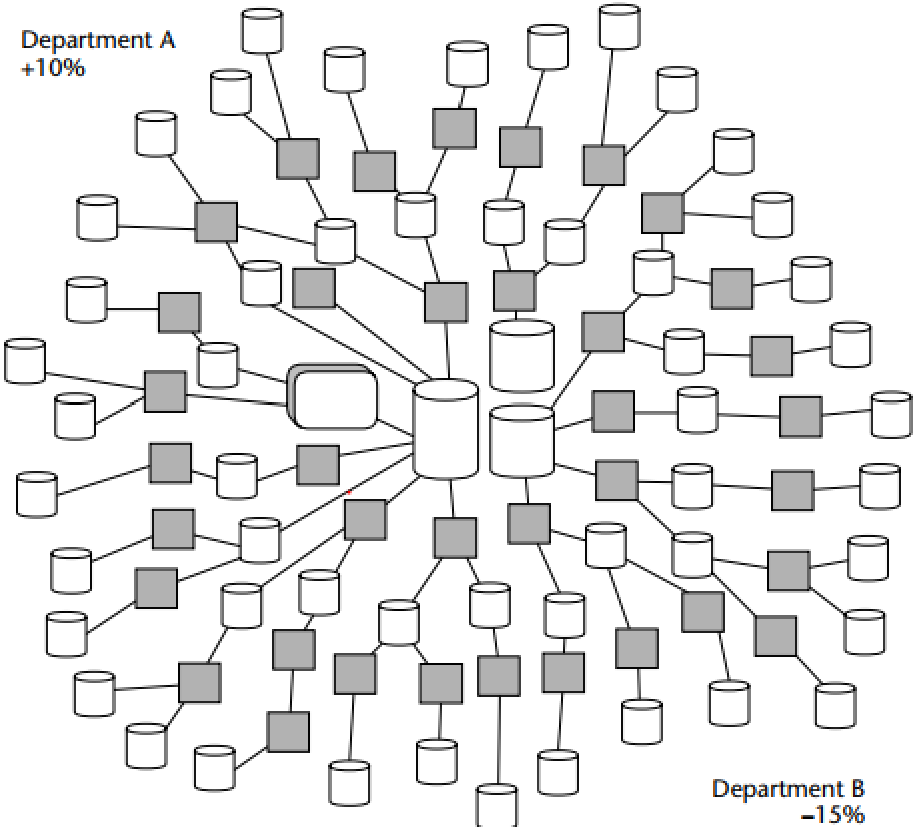
https://www.tableau.com/learn/articles/business-intelligence/bi-business-analytics

BI CHALLENGES

- From business
  - Reports meet the needs of business requirements and different stakeholders
  - Report usability
  - More than expectation
  - Etc.
- From technique and technology
  - BI system development
  - Different levels of BI
  - The support of BI technology
  - Etc.

WHAT'S MORE…

- Lack of credibility
  - Sales says revenue is +10%
  - Finance says revenue is -15%
  - SCM says revenue is +5%
- Reasons
  - No time basis of data
  - The algorithmic differential of data
  - The levels of extraction
  - The problem of external data
  - No common source of data from the beginning
  - Errors of all kinds
  - Etc.

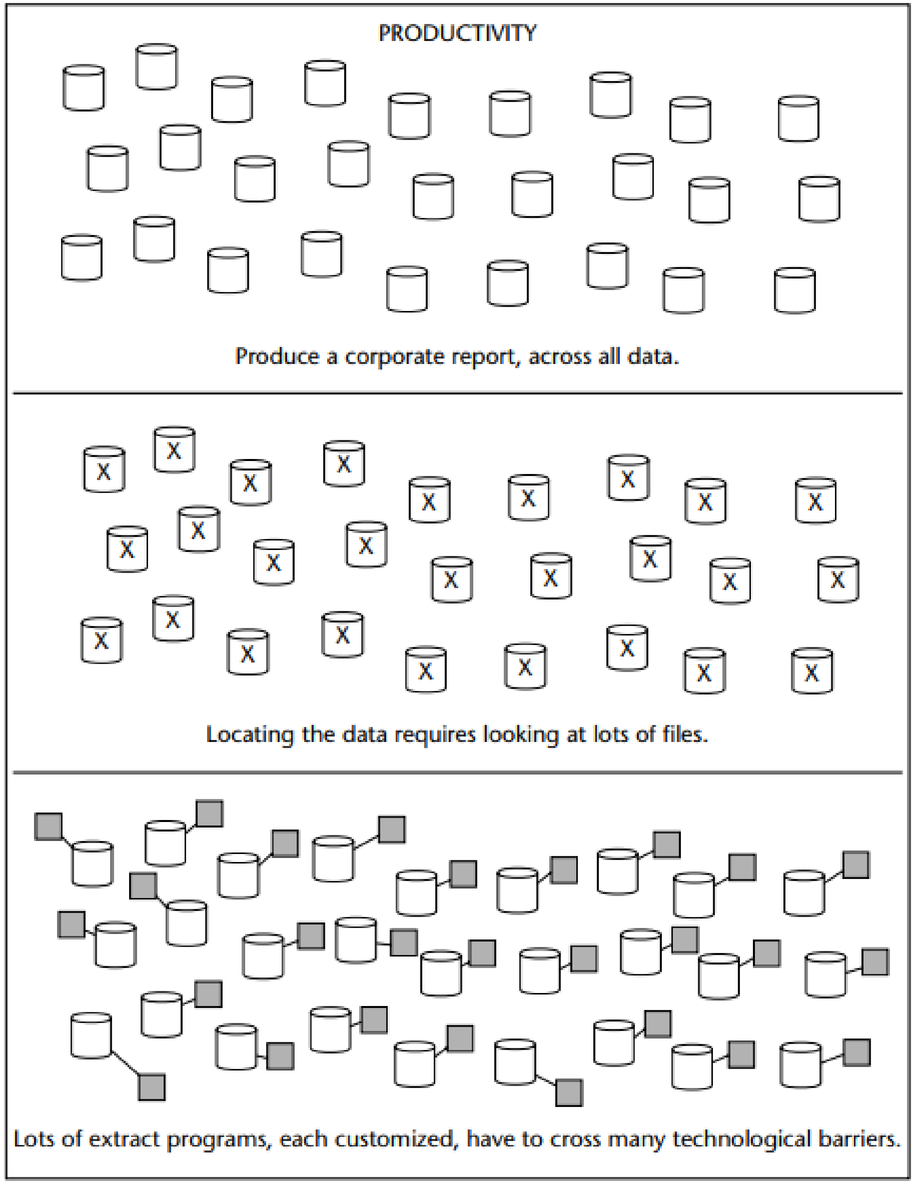


[1]

WHAT'S MORE…

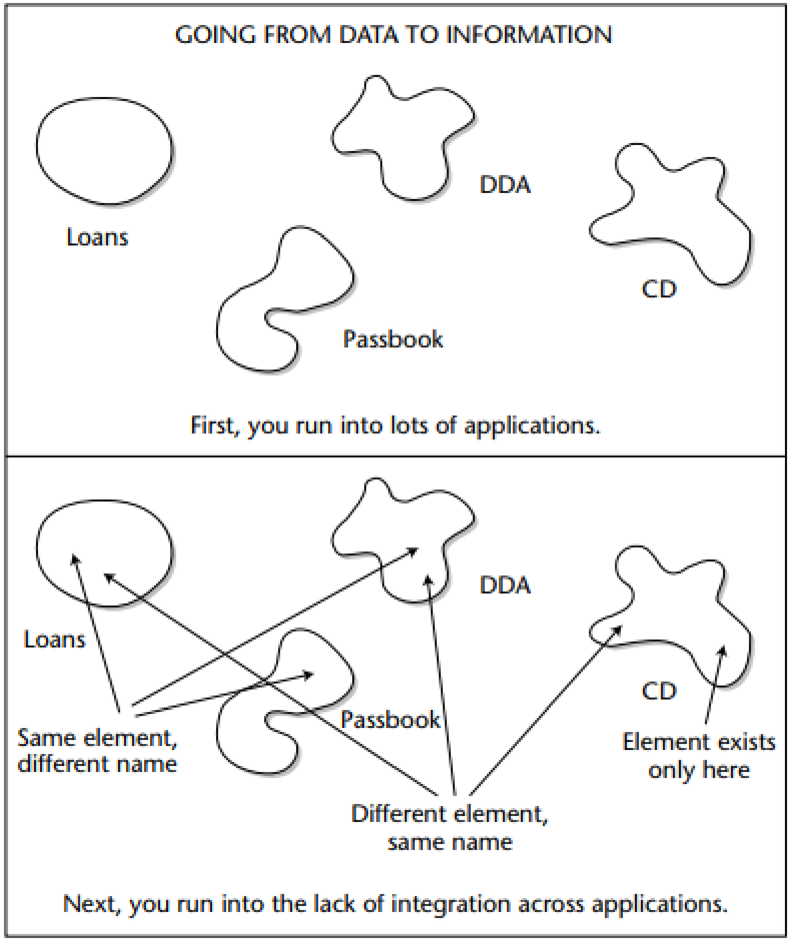
- Problems of productivity
  - Locate and analyze the data for the report.
  - Compile the data for the report.
  - Get programmer/analyst resources to accomplish these two tasks.
- Reasons
  - Big data (3Vs)
  - Technology variance
  - Reusability
  - Unknown future demands
  - Etc.

[1]


PRODUCTIVITY
Produce a corporate report, across all data.
X
Locating the data requires looking at lots of files.
Lots of extract programs, each customized, have to cross many technological barriers.

WHAT'S MORE…

- Problems of obtaining information
  - Data is unavailable
  - Data is not integrated
- Reasons
  - No data history
  - Unforeseen data needs
  - Unintegrated legacy applications
  - Etc.


GOING FROM DATA TO INFORMATION
Loans
DDA
CD
Passbook
First, you run into lots of applications.
Loans
DDA
CD
Passbook
Same element, different name
Element exists only here
Different element, same name
Next, you run into the lack of integration across applications.

[1]

ARCHITECTED ENVIRONMENT

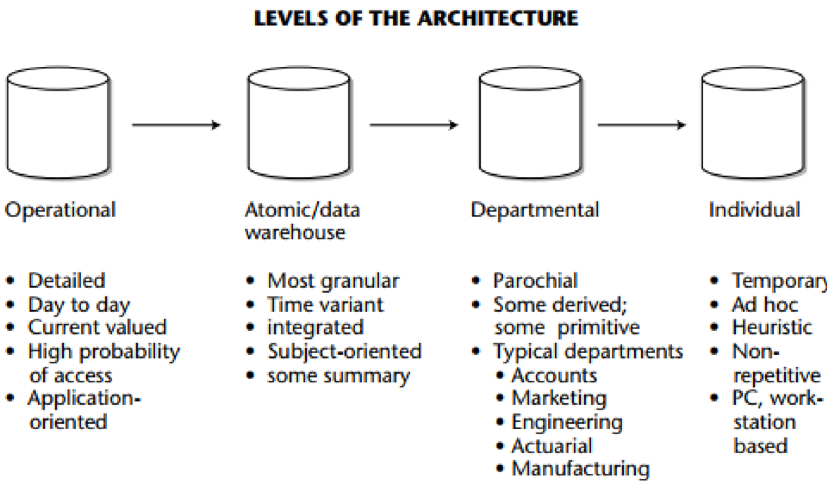
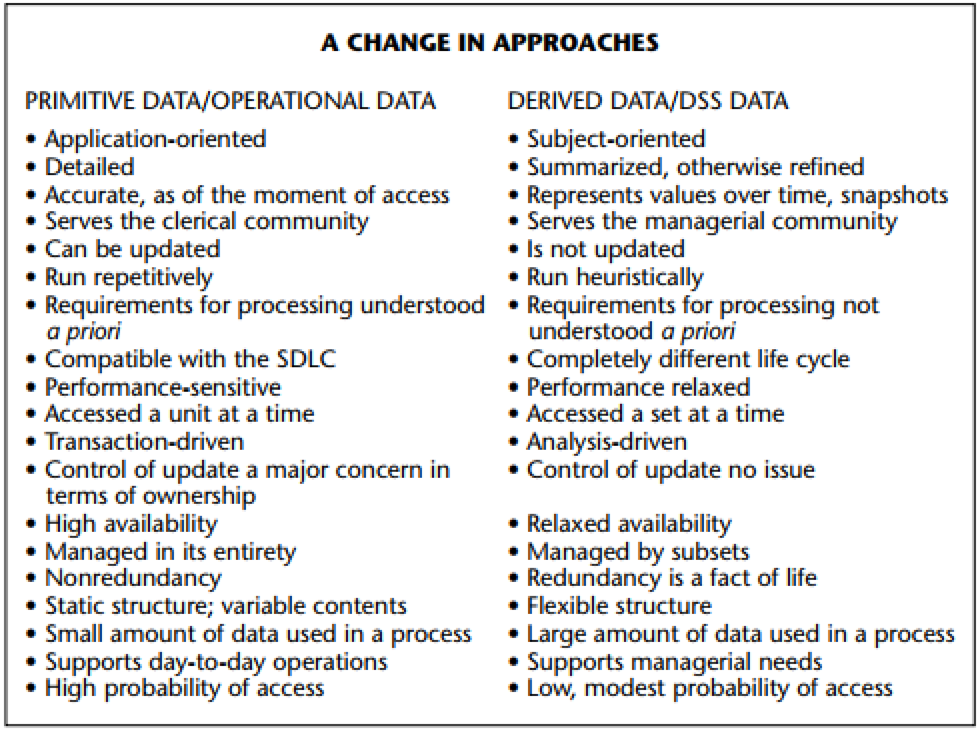


Figure 1-10 Although it is not apparent at first glance, there is very little redundancy data across the architected environment.



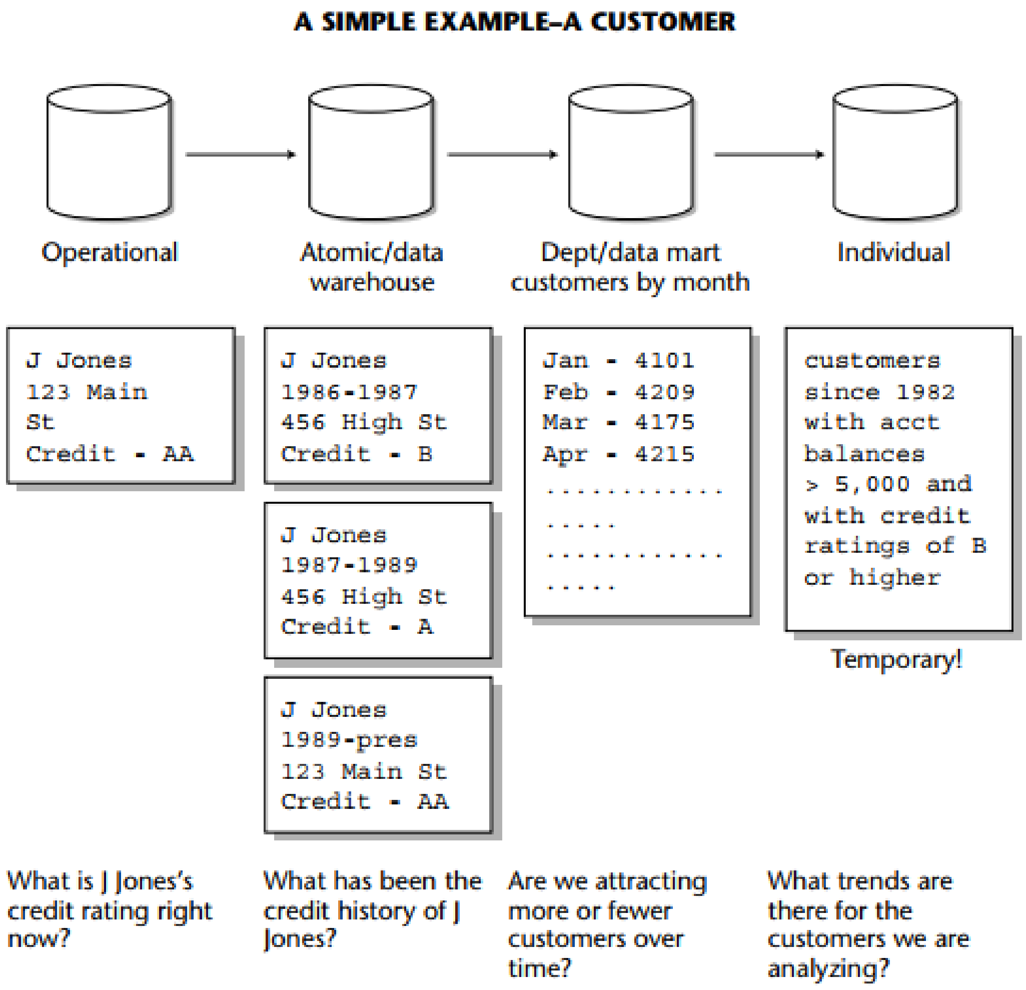
A CHANGE IN APPROACHES

PRIMITIVE DATA/OPERATIONAL DATA	DERIVED DATA/DSS DATA
• Application-oriented	• Subject-oriented
• Detailed	• Summarized, otherwise refined
• Accurate, as of the moment of access	• Represents values over time, snapshots
• Serves the clerical community	• Serves the managerial community
• Can be updated	• Is not updated
• Run repetitively	• Run heuristically
• Requirements for processing understood *a priori*	• Requirements for processing not understood *a priori*
• Compatible with the SDLC	• Completely different life cycle
• Performance-sensitive	• Performance relaxed
• Accessed a unit at a time	• Accessed a set at a time
• Transaction-driven	• Analysis-driven
• Control of update a major concern in terms of ownership	• Control of update no issue
• High availability	• Relaxed availability
• Managed in its entirety	• Managed by subsets
• Nonredundancy	• Redundancy is a fact of life
• Static structure; variable contents	• Flexible structure
• Small amount of data used in a process	• Large amount of data used in a process
• Supports day-to-day operations	• Supports managerial needs
• High probability of access	• Low, modest probability of access

Figure 1-9 Differences between primitive and derived data.

[1]

AN EXAMPLE



A SIMPLE EXAMPLE–A CUSTOMER

Operational

J Jones
123 Main
St
Credit - AA

What is J Jones's credit rating right now?

Atomic/data warehouse

J Jones
1986-1987
456 High St
Credit - B

J Jones
1987-1989
456 High St
Credit - A

J Jones
1989-pres
123 Main St
Credit - AA

What has been the credit history of J Jones?

Dept/data mart customers by month

Jan - 4101
Feb - 4209
Mar - 4175
Apr - 4215
....
.....
....
.....

Are we attracting more or fewer customers over time?

Individual

customers
since 1982
with acct
balances
> 5,000 and
with credit
ratings of B
or higher

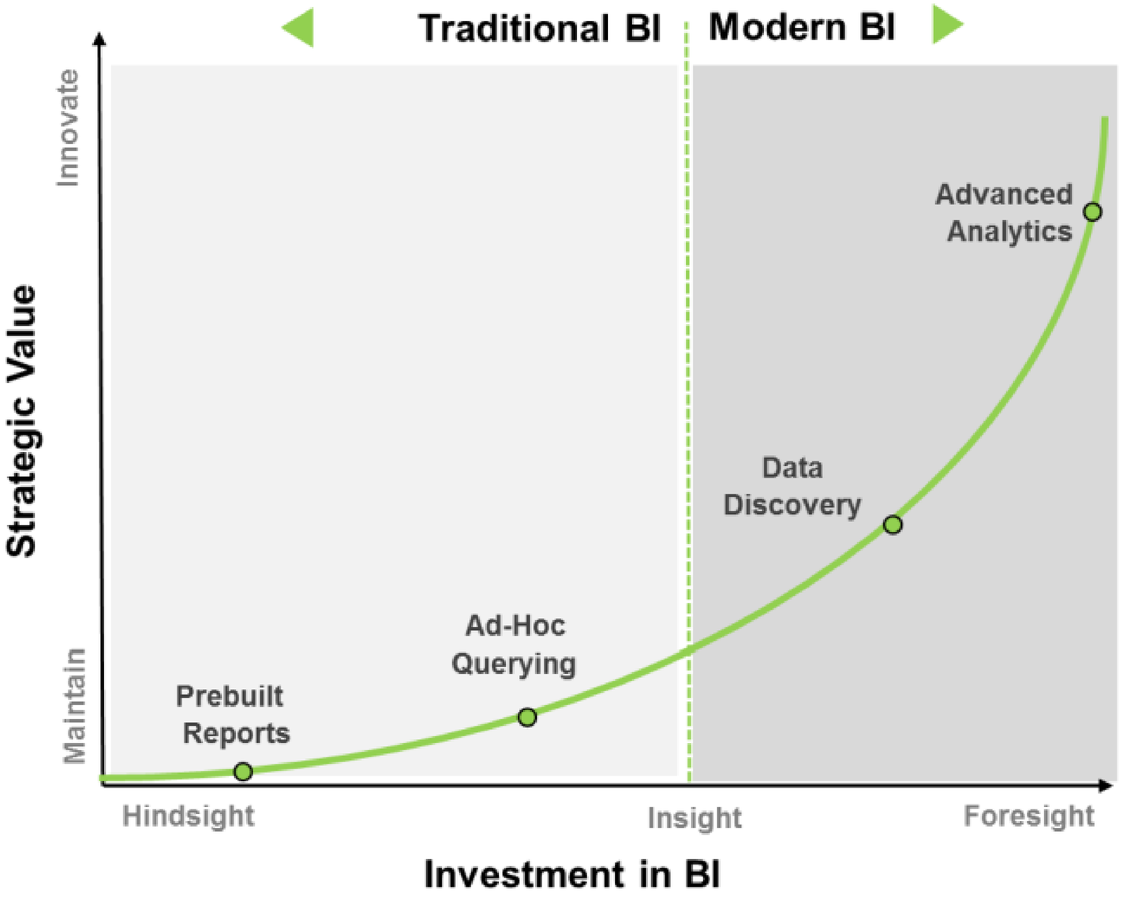
Temporary!

What trends are there for the customers we are analyzing?

Figure 1-11 The kinds of queries for which the different levels of data can be used.

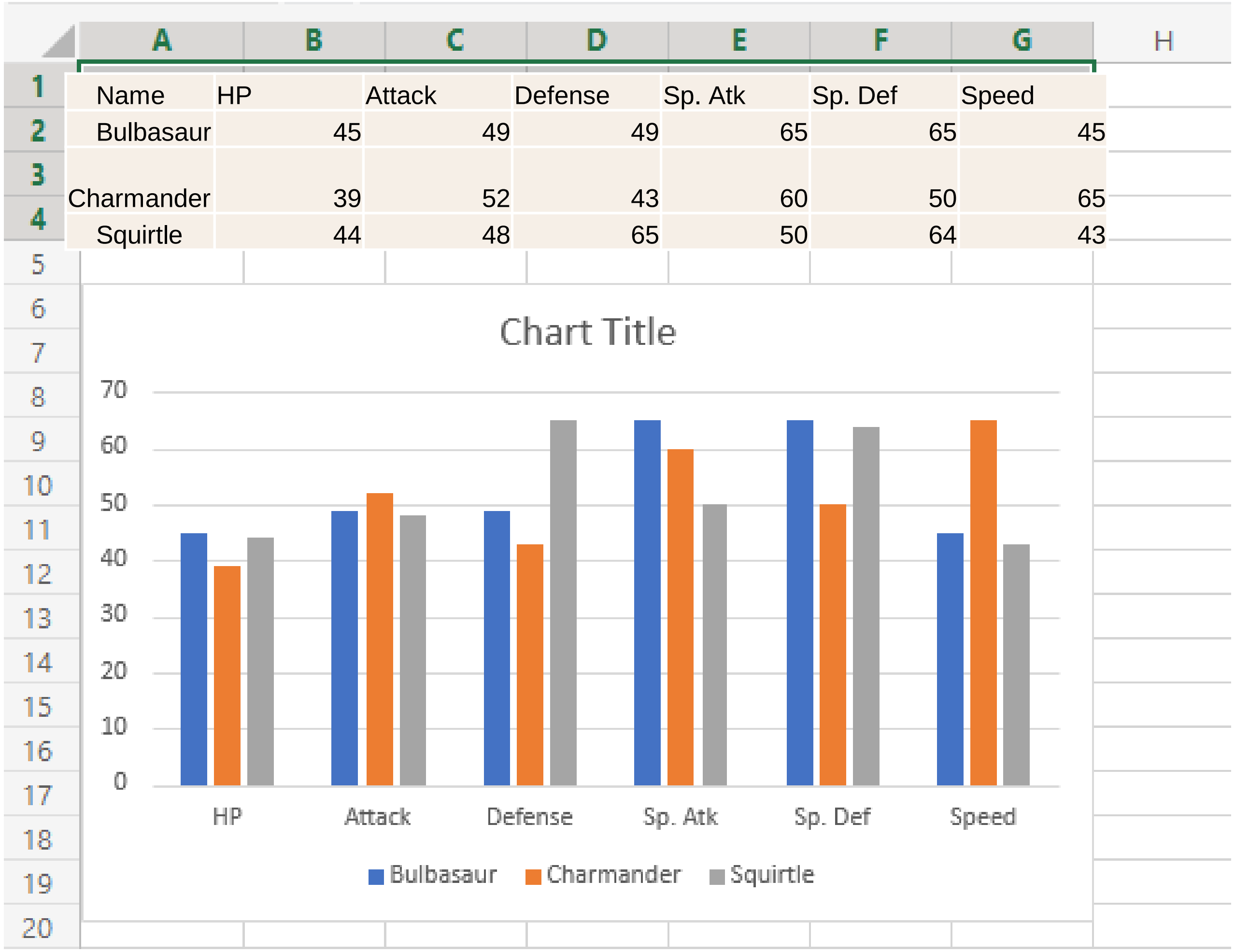
[1]

TRADITIONAL TO MODERN BI (1/4)



Deloitte: Modern Business Intelligence, The Path to Big Data Analytics, 2018.

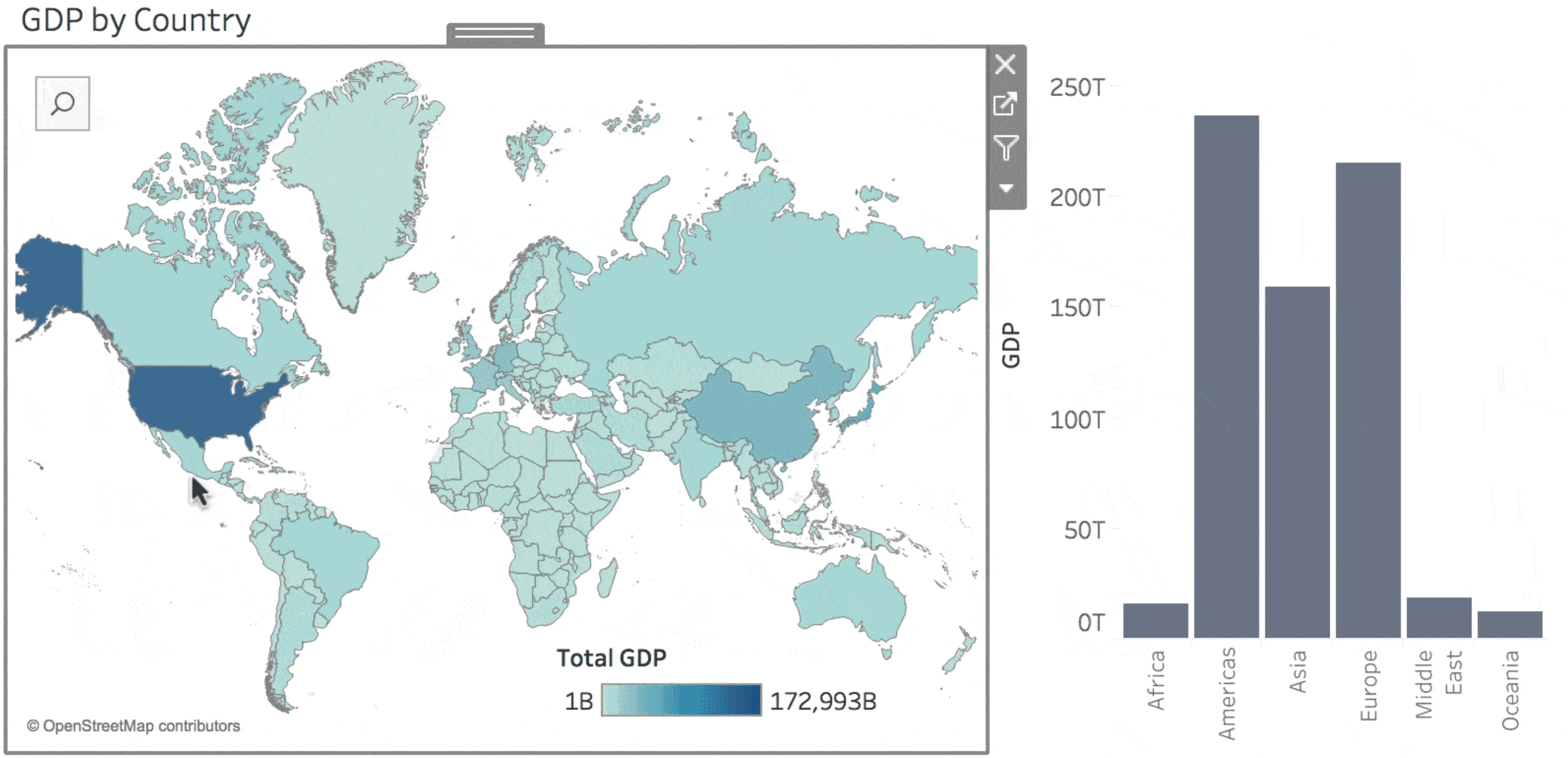
TRADITIONAL TO MODERN BI (2/4)



Name	HP	Attack	Defense	Sp. Atk	Sp. Def	Speed
Bulbasaur	45	49	49	65	65	45
Charmander	39	52	43	60	50	65
Squirtle	44	48	65	50	64	43

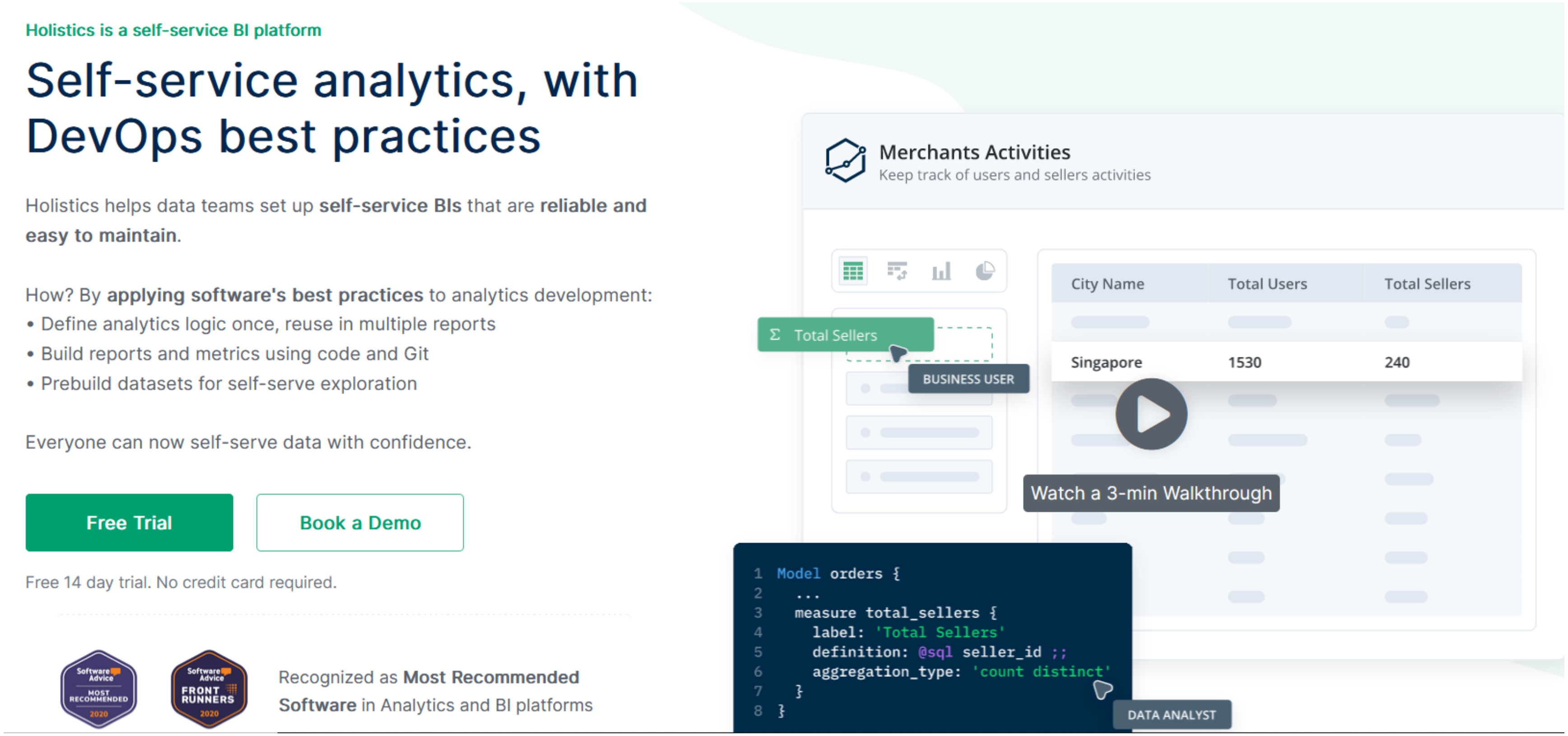
https://www.w3schools.com/excel/img_excel_chart_intro_8_5.png

TRADITIONAL TO MODERN BI (3/4)



https://www.tableau.com/learn/articles/business-intelligence

TRADITIONAL TO MODERN BI (3/4)



Holistics is a self-service BI platform

Self-service analytics, with DevOps best practices

Holistics helps data teams set up **self-service BIs** that are **reliable and easy to maintain**.

How? By **applying software's best practices** to analytics development:

- Define analytics logic once, reuse in multiple reports
- Build reports and metrics using code and Git
- Prebuild datasets for self-serve exploration

Everyone can now self-serve data with confidence.

Free 14 day trial. No credit card required.

Software Advice MOST RECOMMENDED 2020
Software Advice FRONT RUNNERS 2020

Recognized as **Most Recommended Software** in Analytics and BI platforms

Merchants Activities

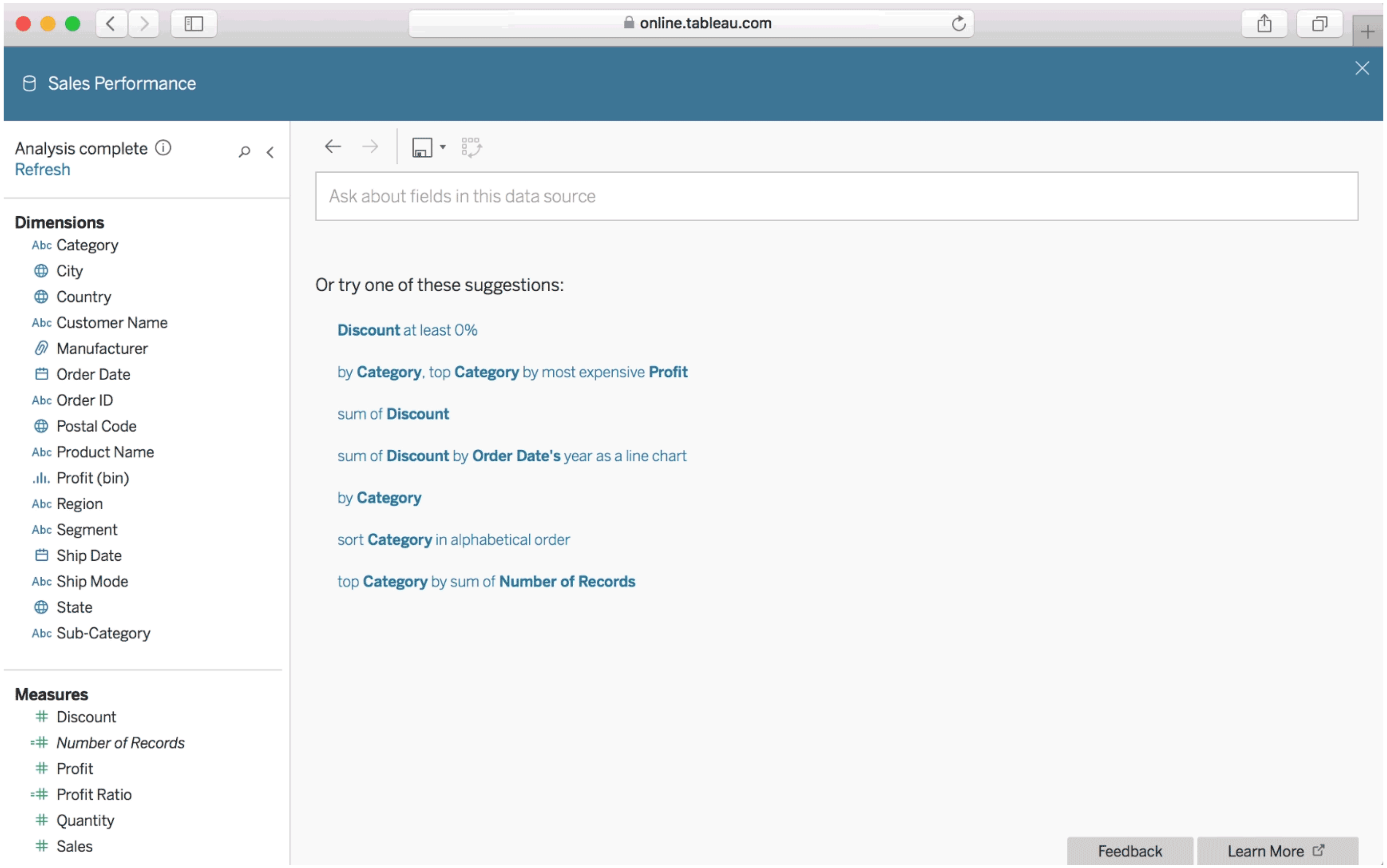
Keep track of users and sellers activities

City Name	Total Users	Total Sellers
Singapore	1530	240

```
Model orders {
  ...
  measure total_sellers {
    label: 'Total Sellers'
    definition: @sql seller_id ;;
    aggregation_type: 'count distinct'
  }
}
```

https://www.holistics.io/

TRADITIONAL TO MODERN BI (4/4)



Ask about fields in this data source

Dimensions

- Abc Category
- City
- Country
- Abc Customer Name
- Manufacturer
- Order Date
- Abc Order ID
- Postal Code
- Abc Product Name
- Profit (bin)
- Abc Region
- Abc Segment
- Ship Date
- Abc Ship Mode
- State
- Abc Sub-Category

Measures

- Discount
- *Number of Records*
- Profit
- Profit Ratio
- Quantity
- Sales

Or try one of these suggestions:

- **Discount** at least 0%
- by **Category**, top **Category** by most expensive **Profit**
- sum of **Discount**
- sum of **Discount** by **Order Date's** year as a line chart
- by **Category**
- sort **Category** in alphabetical order
- top **Category** by sum of **Number of Records**

https://www.tableau.com/learn/articles/business-intelligence

BI PROCESSES AND ACTIVITIES

- Reporting and visualization
- Statistical analysis
- Data mining
- Performance metrics and benchmarking
- Visual storytelling
- Data preparation
- Querying
- Etc.

BI STRATEGY BY TABLEAU

1. Know your business strategy and goals.
2. Identify key stakeholders.
3. Choose a sponsor from your key stakeholders.
4. Choose your BI platform and tools.
5. Create a BI team.
6. Define your scope.
7. Prepare your data infrastructure.
8. Define your goals and roadmap.

https://www.tableau.com/learn/articles/business-intelligence

DISCUSSION

BUSINESS ANALYTICS
VS.
DATA ANALYTICS

DISCUSSION

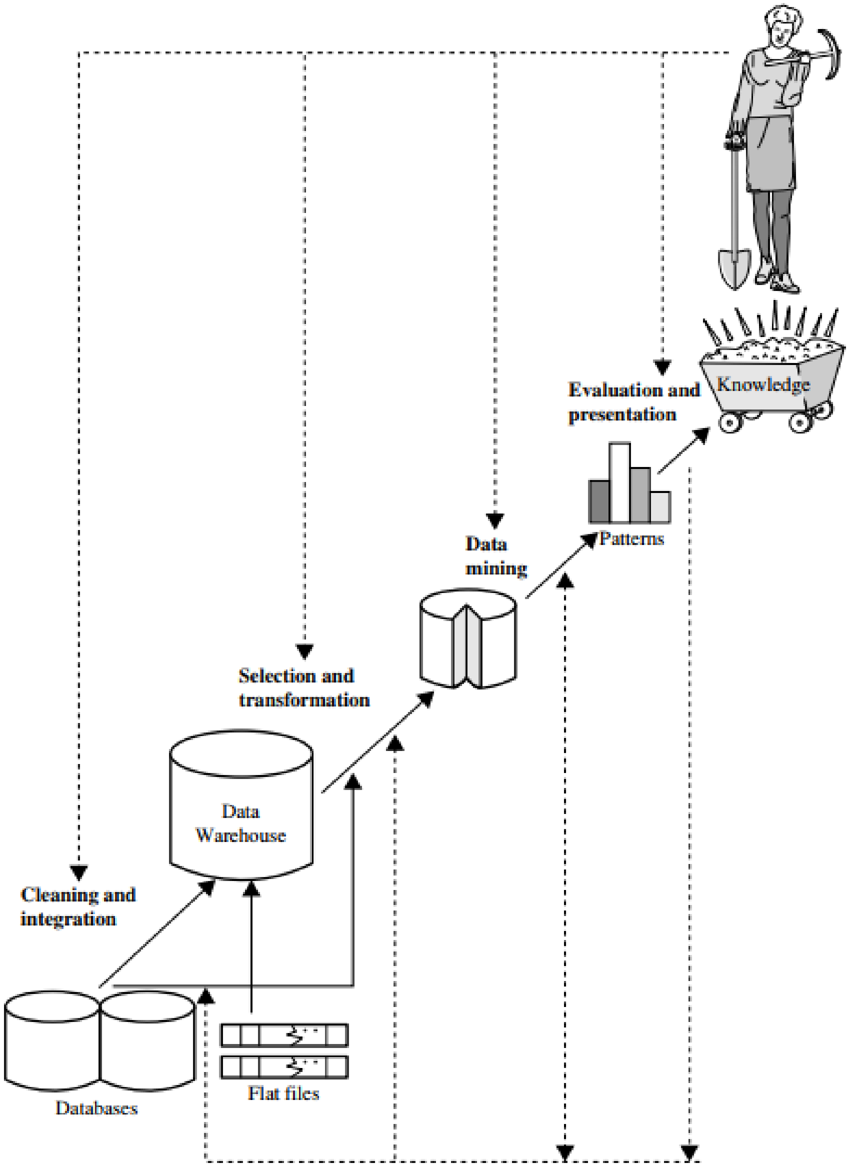
BI PROS AND CONS

DISCUSSION

BI APPLICATION
(E.g., PowerBI, Tableau, Looker Studio, Holistics)

DATA MINING OVERVIEW

KNOWLEDGE DISCOVERY



[2]

Figure 1.4 Data mining as a step in the process of knowledge discovery.

NORMALIZATION VS. DE-NORMALIZATION

- **Normalization** is used to decompose one table data into different sub-tables data
- **Denormalization** is used to combine multiple table data into one

OLTP vs. OLAP

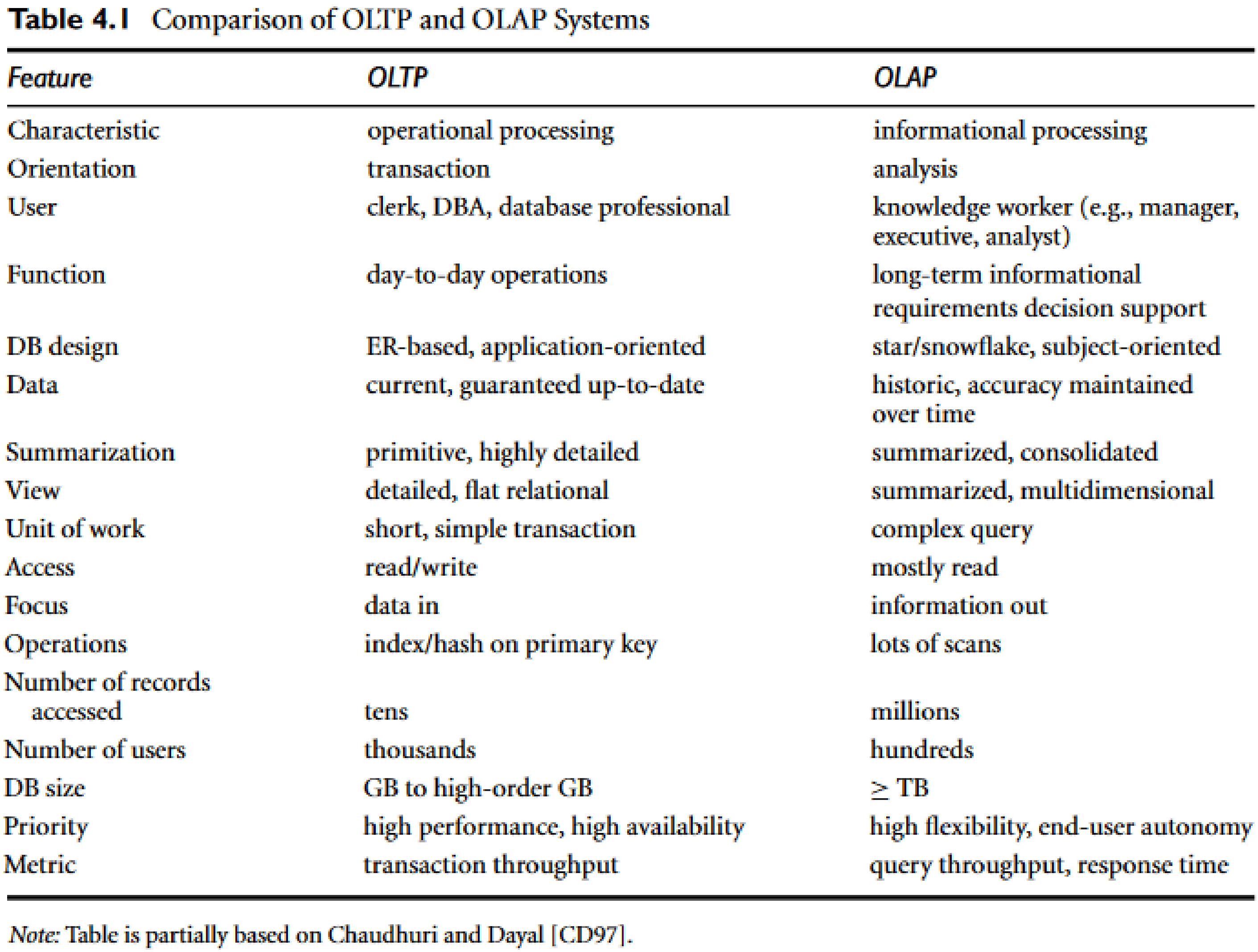


Table 4.1 Comparison of OLTP and OLAP Systems

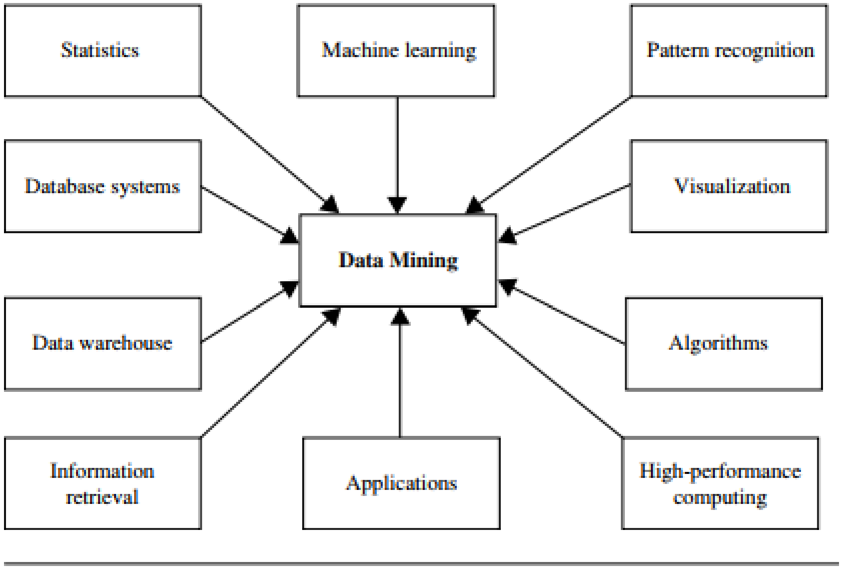
Feature	*OLTP*	*OLAP*
Characteristic	operational processing	informational processing
Orientation	transaction	analysis
User	clerk, DBA, database professional	knowledge worker (e.g., manager, executive, analyst)
Function	day-to-day operations	long-term informational requirements decision support
DB design	ER-based, application-oriented	star/snowflake, subject-oriented
Data	current, guaranteed up-to-date	historic, accuracy maintained over time
Summarization	primitive, highly detailed	summarized, consolidated
View	detailed, flat relational	summarized, multidimensional
Unit of work	short, simple transaction	complex query
Access	read/write	mostly read
Focus	data in	information out
Operations	index/hash on primary key	lots of scans
Number of records accessed	tens	millions
Number of users	thousands	hundreds
DB size	GB to high-order GB	≥ TB
Priority	high performance, high availability	high flexibility, end-user autonomy
Metric	transaction throughput	query throughput, response time

Note: Table is partially based on Chaudhuri and Dayal [CD97].

[8]

DATA MINING

- Data mining is the process of discovering interesting patterns and knowledge from large amounts of data.



[2] **Figure 1.11** Data mining adopts techniques from many domains.

DATA MINING FUNCTIONALITY

- Descriptive tasks
  - characterize properties of the data in a target data set
- Predictive tasks
  - perform induction on the current data in order to make predictions

KINDS OF DATA TO BE MINED

- Database data
- Data warehouse data
- Transactional data
- Stream data
- Sequence data
- Graph data
- Spatial data
- Text data
- Multimedia data
- Etc.

KINDS OF PATTERNS TO BE MINED

- Class description
- Frequent patterns
- Associations
- Classification and regression
- Cluster analysis
- Outlier analysis

DATA CHARACTERIZATION

- It summarizes the data of the class under study (often called the target class) in general terms
- E.g.,
  - Summarizing the characteristics of customers who spend more than $5000 a year at AllElectronics.
  - The result is a general profile of these customers, such as that they are 40 to 50 years old, employed, and have excellent credit ratings.
  - The data mining system should allow the customer relationship manager to drill down on any dimension, such as on occupation to view these customers according to their type of employment.

DATA DISCRIMINATION

- It compares the target class with one or a set of comparative classes (often called the contrasting classes)
- E.g.,
  - Comparing two groups of customers—those who shop for computer products regularly (e.g., more than twice a month) and those who rarely shop for such products (e.g., less than three times a year).
  - The resulting description provides a general comparative profile of these customers, such as that 80% of the customers who frequently purchase computer products are between 20 and 40 years old and have a university education, whereas 60% of the customers who infrequently buy such products are either seniors or youths, and have no university degree. Drilling down on a dimension like occupation, or adding a new dimension like income level, may help to find even more discriminative features between the two classes.

FREQUENT PATTERNS

- They are patterns that occur frequently in data
  - Frequent itemsets: a set of items that often appear together in a transactional data set
  - Frequent subsequences: a frequently occurring subsequence
  - Frequent substructures: a frequently occurring structural forms (e.g., graphs, trees, or lattices) that may be combined with itemsets or subsequences.

ASSOCIATIONS

- Mining frequent patterns leads to the discovery of interesting associations within data.
- E.g.,
  - *buys(X,* “*computer*”*)* → *buys(X,* “*software*”*)* with *support* = 1% and *confidence* = 50%
  - *age(X,* “20..29”) and *income(X,* “40K..49K”*)* → *buys(X,* “laptop”) *with support* = 2% and *confidence* = 60%

CLASSIFICATION

- Classification is the process of finding a model (or function) that describes and distinguishes data classes or concepts.

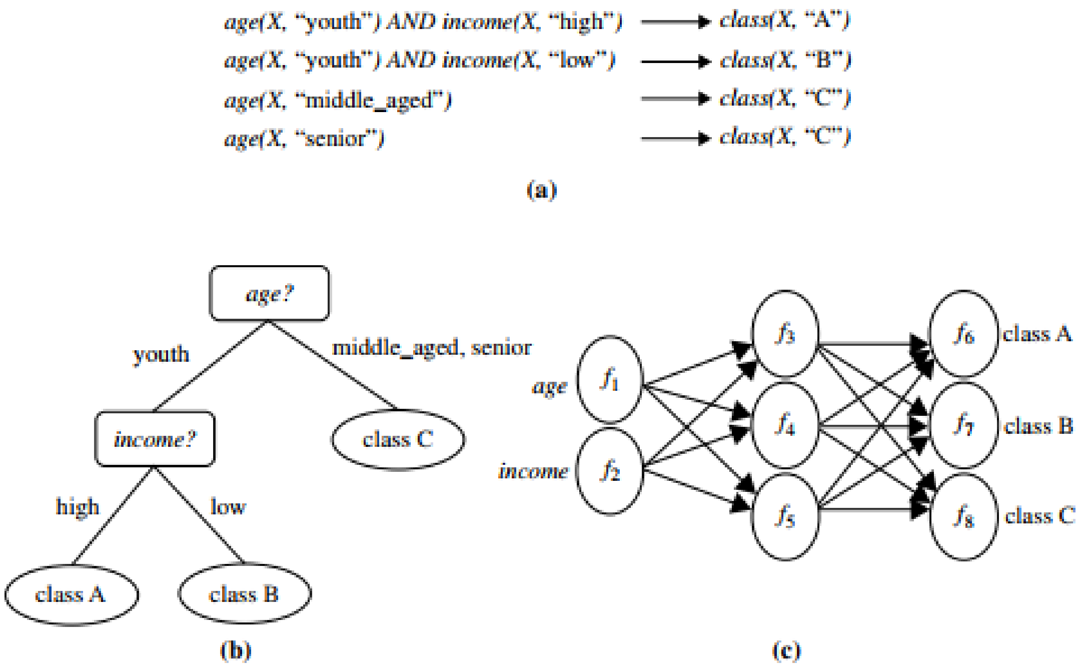
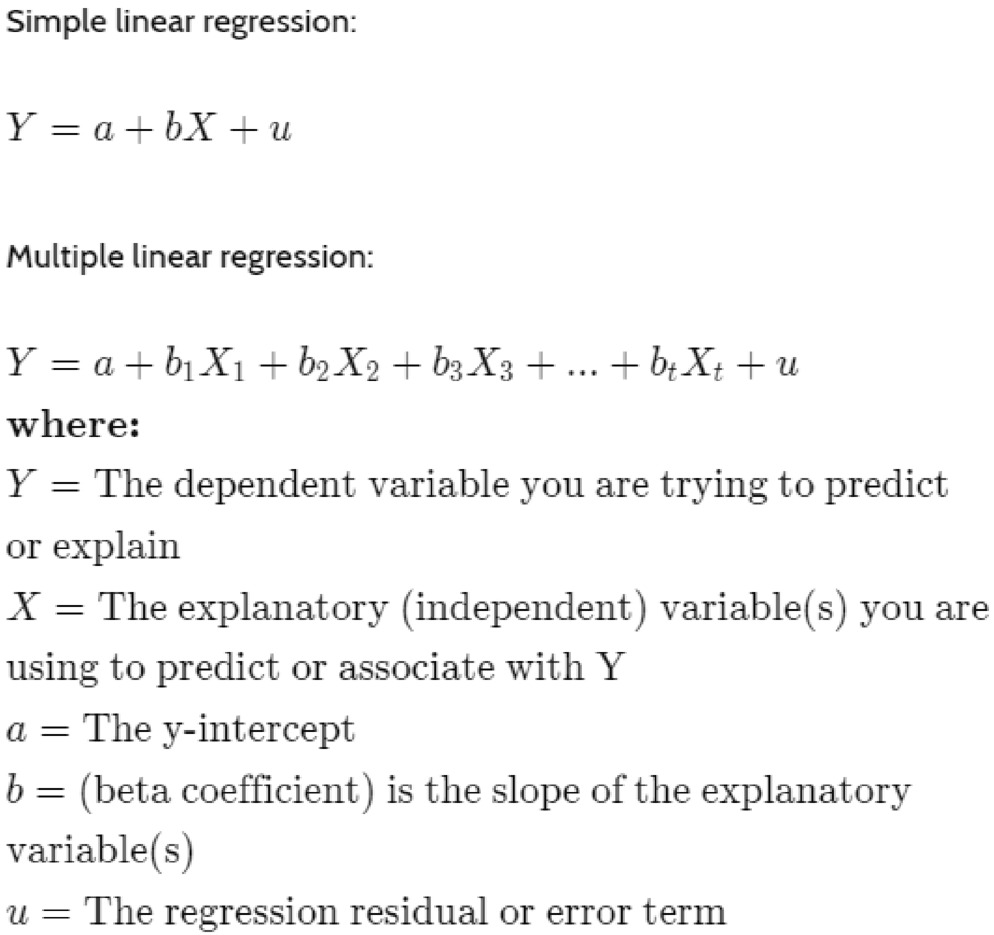

age(X, "youth") AND income(X, "high") → class(X, "A")
age(X, "youth") AND income(X, "low") → class(X, "B")
age(X, "middle_aged") → class(X, "C")
age(X, "senior") → class(X, "C")
(a)
age?
youth
middle_aged, senior
income?
class C
high
low
class A
class B
(b)
age
income
f_1
f_2
f_3
f_4
f_5
f_6
f_7
f_8
class A
class B
class C
(c)

Figure 1.9 A classification model can be represented in various forms: (a) IF-THEN rules, (b) a decision tree, or (c) a neural network.

[2]

REGRESSION

- Regression models continuous-valued functions. That is, regression is used to predict missing or unavailable numerical data values rather than (discrete) class labels.



Simple linear regression:

$$Y = a + bX + u$$

Multiple linear regression:

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3 + ... + b_tX_t + u$$

where:

Y = The dependent variable you are trying to predict or explain

X = The explanatory (independent) variable(s) you are using to predict or associate with Y

a = The y-intercept

b = (beta coefficient) is the slope of the explanatory variable(s)

u = The regression residual or error term

Internet

CLUSTER ANALYSIS

- Clusters of objects are formed so that objects within a cluster have high similarity in comparison to one another, but are rather dissimilar to objects in other clusters.
- Clustering analyzes data objects without consulting class labels.
- E.g., interests based on locations

OUTLIER ANALYSIS

- Objects that do not comply with the general behavior or model of the data. These data objects are outliers.
- E.g., fraudulent usage of credit cards

DISCUSSION

ARE ALL PATTERNS INTERESTING?

DISCUSSION

DATA MINING VS. MACHINE LEARNING?

TO BE CONTINUED…

- Lecture 2: Data pre-processing

QUESTIONS AND ANSWERS

Picture from: http://philadelphiasculpturegym.blogspot.com/2013/09/save-date-free-talk-and-q-on-affordable.html

REFERENCES

[1] Tobias Zwingmann, AI-Powered Business Intelligence, Kindle Edition, O'reilly Press, 2022

[2] Jiawei Han, Micheline Kamber, “Data Mining: Concepts and Techniques”, Third Edition, Morgan Kaufmann Publishers, 2012.

[3] Jeen Su Lim, John Heinrichs, “Digital Business Intelligence Management with Big Data Analytics" Kindle Edition, O'reilly Press, 2021.

[4] David L. Olson, Dursun Delen, “Advanced Data Mining Techniques”, Springer-Verlag, 2008.

[5] Brian Larson, “Delivering Business Intelligence with Microsoft SQL Server 2016”, McGraw-Hill Education; 4 edition, 2016.

[6] Oracle, “Data Mining Concepts”, 18c, E83730-03, 2018

[7] Oracle, “Data Mining Application Developer’s Guide”, 2013.