VII International conference
"Information Technology and Interactions" (IT&I-2020)
Kyiv, Ukraine

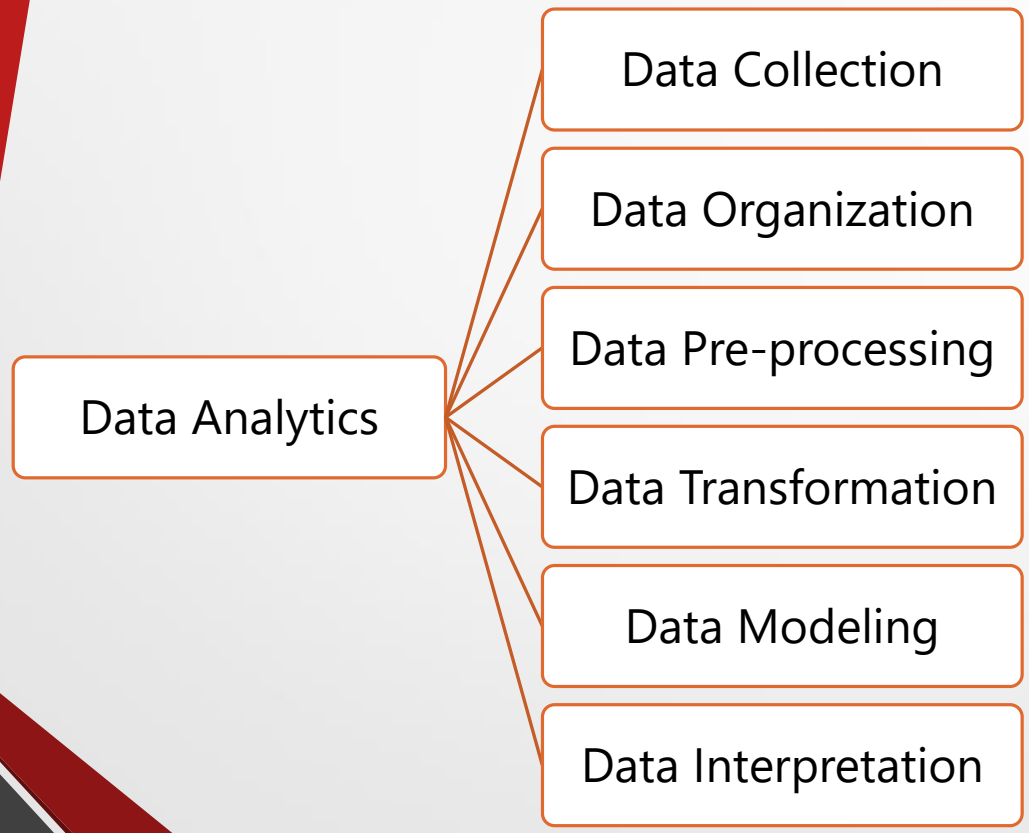
A Business Intelligence Dashboard Design Approach to Improve Data Analytics and Decision Making

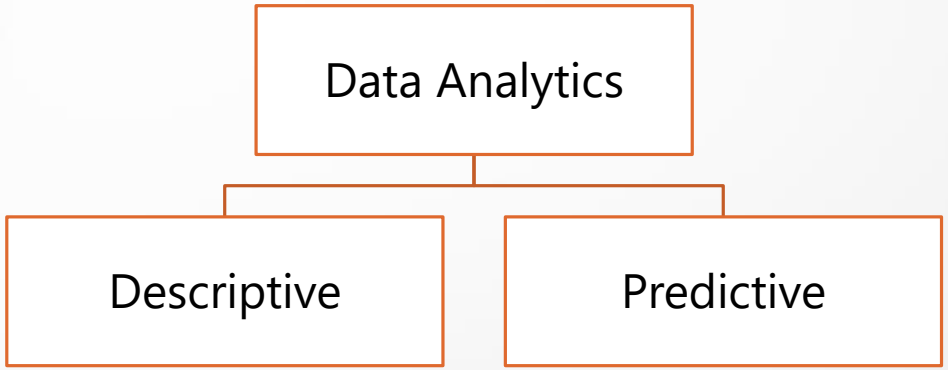
Orlovskyi Dmytro, Andrii Kopp

Presentation Structure

1. Data Analytics and Business Intelligence in Decision Making
2. Data Visualization and Business Intelligence Dashboards
3. Problem Statement
4. Dataset Preparation
5. Dataset Analysis
6. Dashboard Design Process
7. Obtained Results
8. Conclusion and Future Work

1. Data Analytics and Business Intelligence in Decision Making

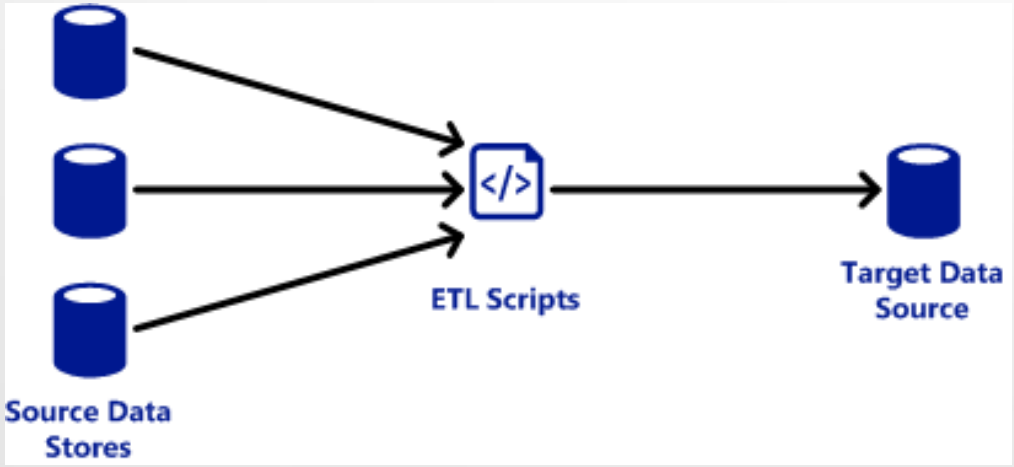




Business Intelligence (BI) is an umbrella term that includes the applications, infrastructure and tools, and best practices that enable access to and analysis of information to improve and optimize decisions and performance (Gartner IT Glossary)

1. Data Analytics and Business Intelligence in Decision Making

- BI Capabilities
  - Data Extraction & Processing
  - Data Storage
  - Data Visualization
- IT Solutions for BI
  - Extract-Transform-Load (ETL)
  - Data Warehouses (DW) and Data Marts (DM)
  - Reports and Dashboards



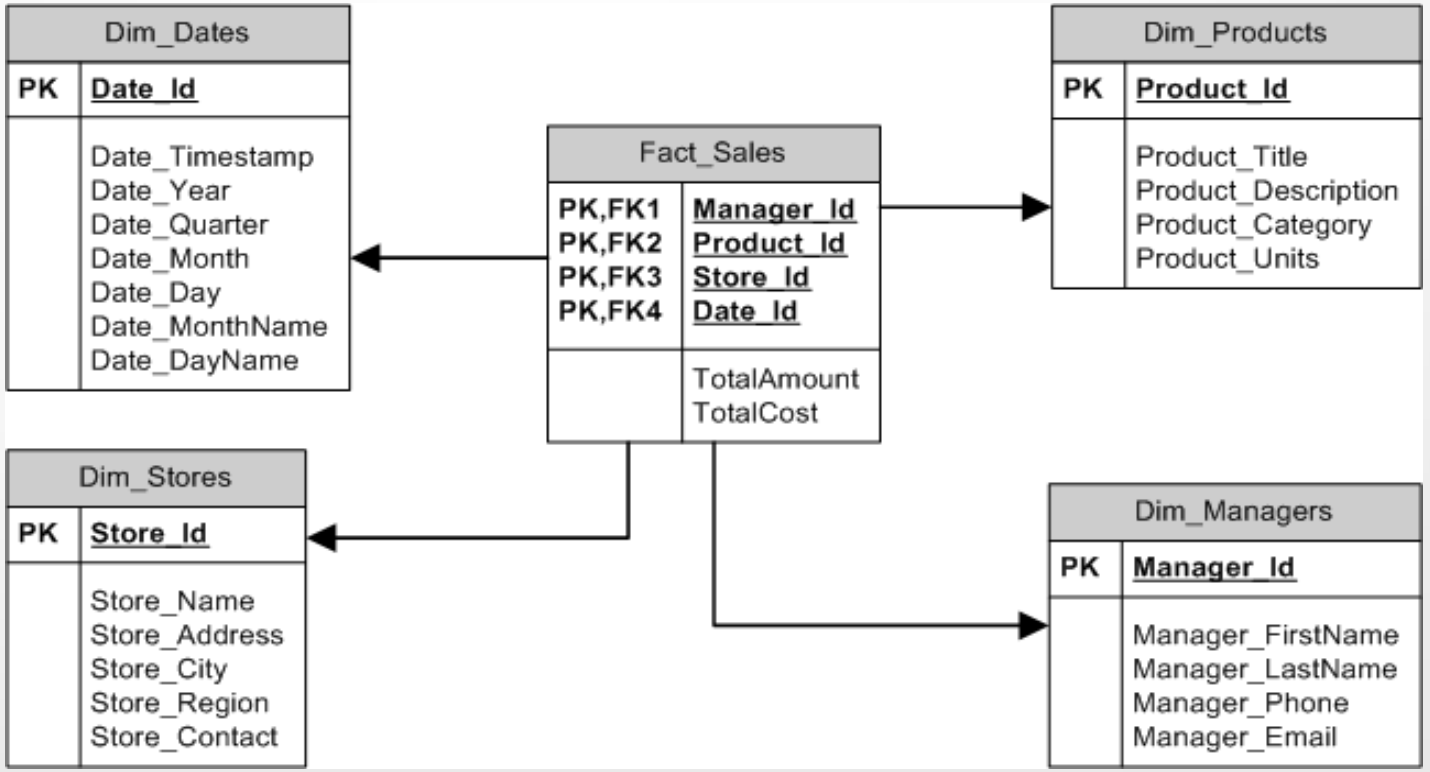
- DM/DW Schemas
  - Star Schema
  - Snowflake Schema

1. Data Analytics and Business Intelligence in Decision Making

Dimensional Modeling Process:

1. Identify a business process
2. Identify a level of detail
3. Identify dimensions
4. Identify facts
5. Build a schema

Sample star schema →



2. Data Visualization and Business Intelligence Dashboards

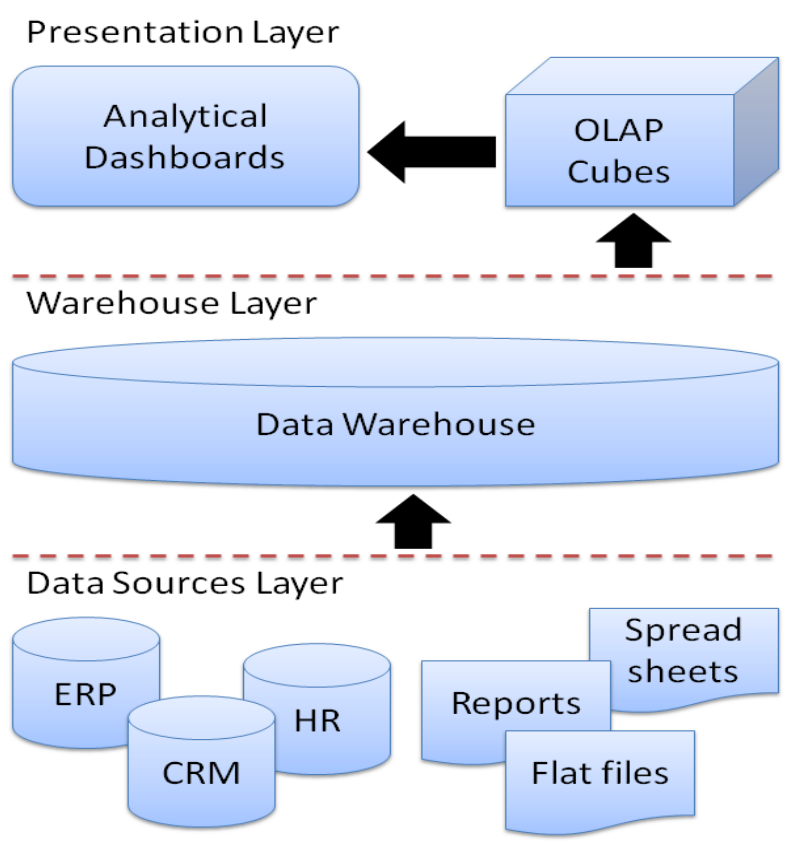
Microsoft Power BI

Qlik

tableau

D3.js

kibana



Dashboards are applications for business users that visualize data by combining charts, graphs, and other data visualizations together

Dashboard Design Principles:

1. Used visualizations should fit the best nature of data
2. Used visualizations should serve their purposes even in case if they are resized in order to be placed into small spaces

2. Data Visualization and Business Intelligence Dashboards

Bar Charts

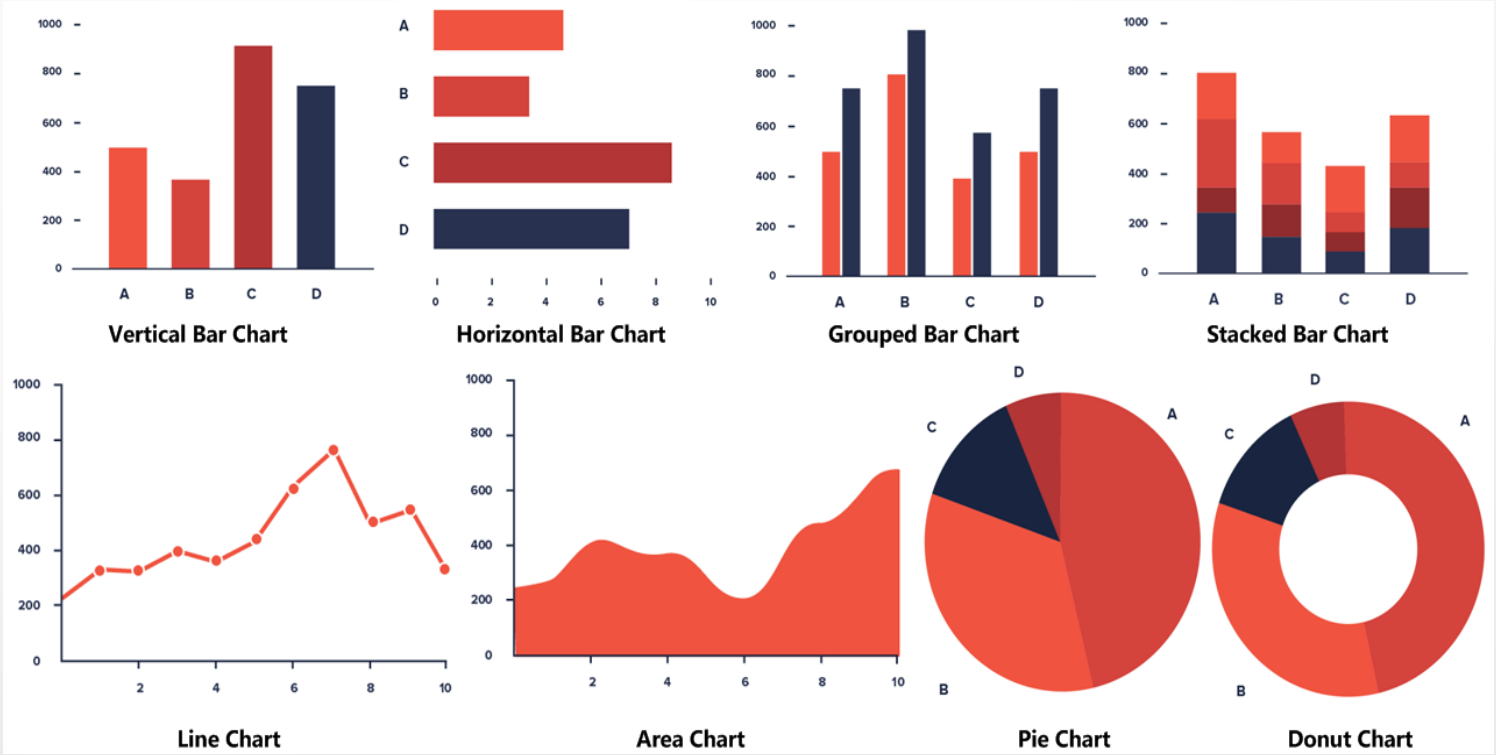
- Comparisons by different measures

Line Charts

- Trends over time by different measures

Pie Charts

- Parts of a whole to compare relative sizes



3. Problem Statement

Dashboard Design Steps:

1. Selection of visualizations (graphs and charts)
2. Placing visualizations in a limited space

Dashboard Design Issues:

1. Inappropriate visualizations do not fit nature of displayed data
2. They may mislead business users and shift their attention to unimportant or wrong things

Purpose: improve Business Intelligence capabilities by providing a dashboard design approach that may be used to suggest appropriate data visualizations

4. Dataset Preparation

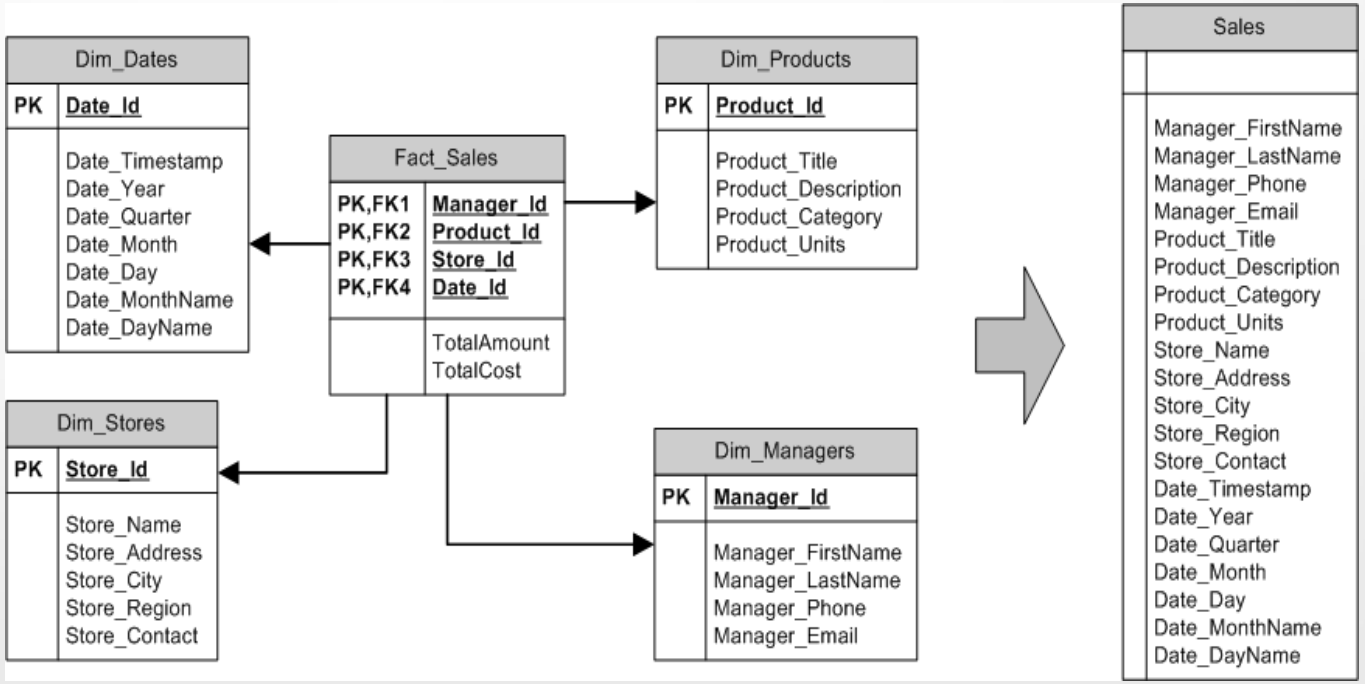
Translate the star schema into flat structure:

1. Join dimension tables and the fact table
2. Include measures and dimension attributes into the result set

$$\{M_j | j = \overline{1,m}\} \bowtie_{D_i^{PK}, i=\overline{1,n}} \{D_i^k | k = \overline{1,q_i}\},$$

$$\{M_j | j = \overline{1,m}\} \cup \bigcup_{i=\overline{1,n}} \{D_i^k | k = \overline{1,q_i}\}.$$

- M_j is the j-th measure of the fact table
- D_i^k is the k-th attribute of i-th dimension table
- m, n, q_i are numbers of attributes



5. Dataset Analysis

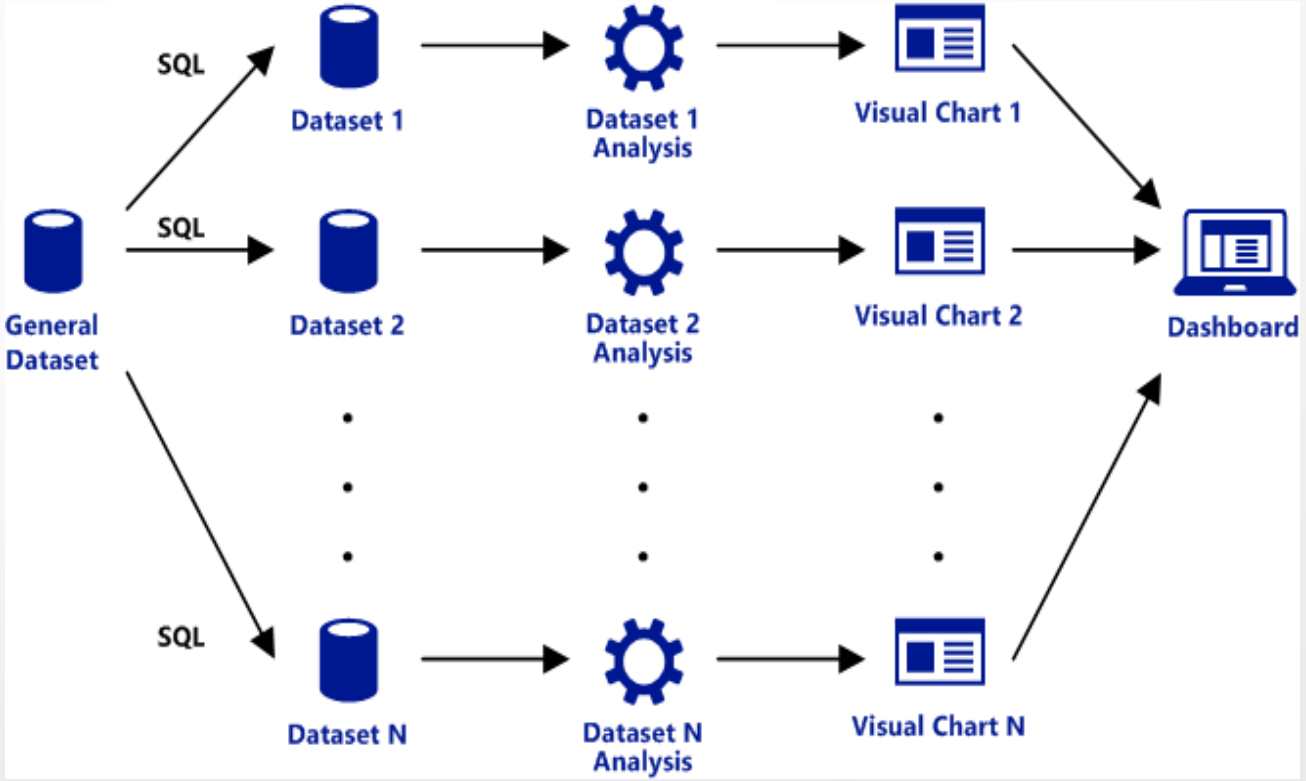
The result set of any SQL query:

$$X \subseteq \bigcup_{i=\overline{1,n}} \{D_i^k | k = \overline{1, q_i}\} \cup Y \subseteq \{M_j | j = \overline{1, m}\},$$

- X is the subset of attributes that may be placed on x-axis
- Y is the subset of attributes that may be placed on y-axis

Visualizations may include:

- Primary and secondary y-axes
- Hierarchical x-axes
- Drill-down and Roll-up operations



5. Dataset Analysis

Proposed function to suggest visualizations for a given dataset DS:

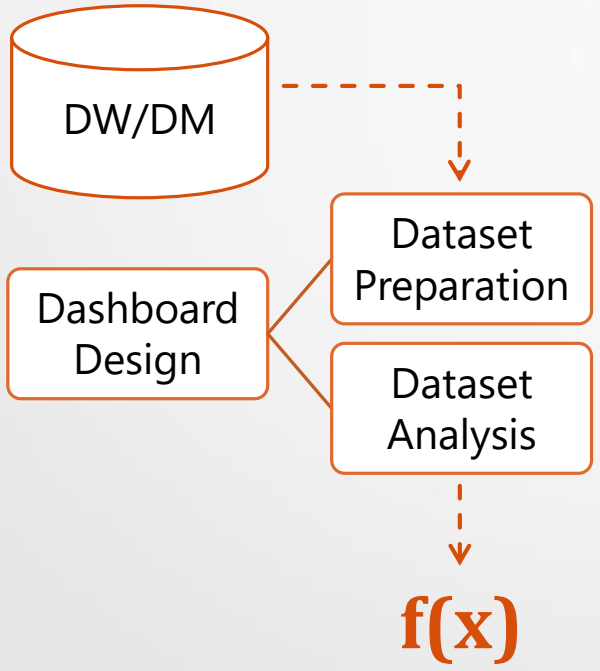
$$Visualization(DS) = \begin{cases} |DS| = 1, & Card \\ 2 \leq |DS| \leq t_1, & Pie \\ t_1 + 1 \leq |DS| \leq t_2, & Bar \\ |DS| > t_2, & Line \end{cases},$$

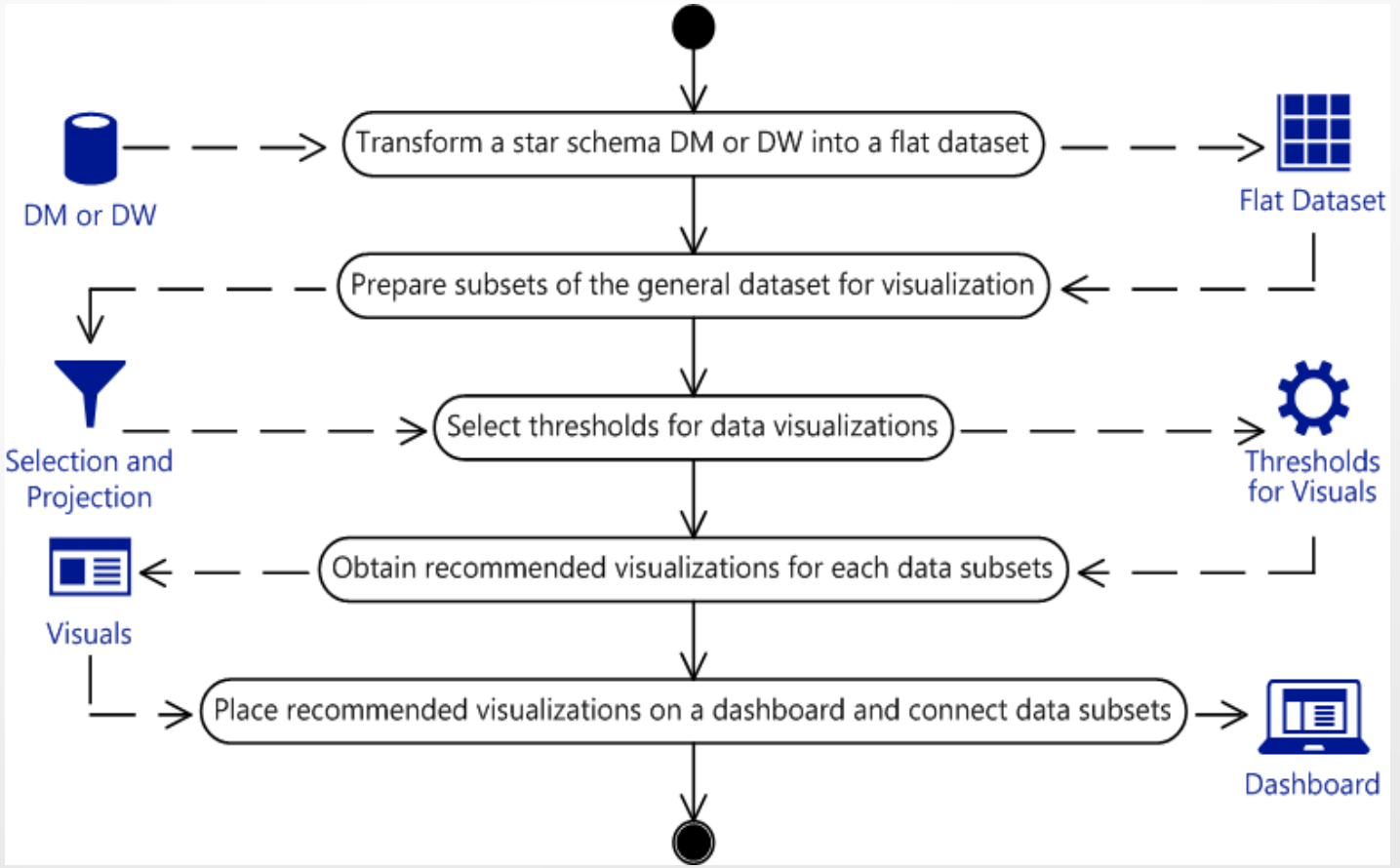
- t_1 is the upper threshold of the dataset size $|DS|$ suitable for a pie chart
- t_2 is the upper threshold of the dataset size $|DS|$ suitable for a bar chart
- $Card$ may be used to display single valued measures (any scalar values)

Source	Pie chart threshold, t_1	Source	Bar chart threshold, t_2
[14, p. 107]	6	*[19]*	10 – 12
[17, p. 68]	4	*[20]*	12
[18, p. 281]	5	*[21]*	15

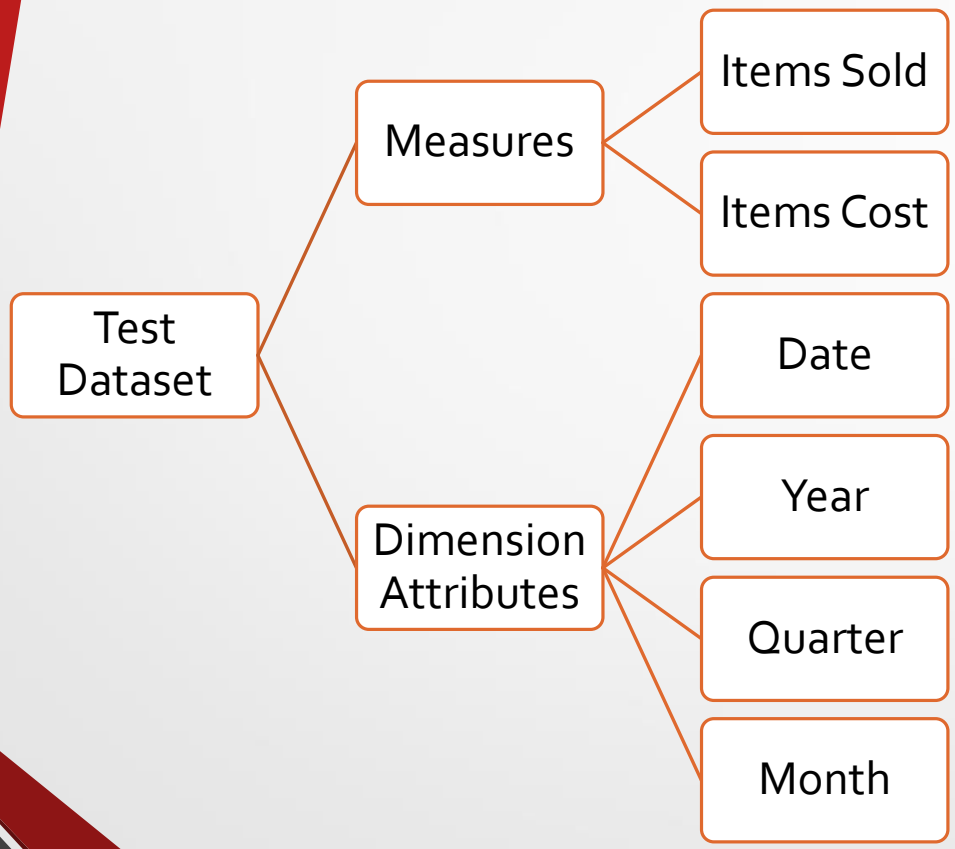
Limitations of pie and bar charts are caused by their usage as comparative tools, while line charts are used to display trends

6. Dashboard Design Process





7. Obtained Results



1. Average revenue by quarter

```
SELECT Quarter, AVG([Items Cost]) AS [Avg Revenue]
FROM Sales
GROUP BY Quarter;
```

2. Lowest and highest sales by month

```
SELECT Month, MIN([Items Sold]) AS [Min Sales], MAX([Items Sold]) AS [Max Sales]
FROM Sales
GROUP BY Month;
```

3. Revenue by date

```
SELECT Date, [Items Cost] AS Revenue
FROM Sales;
```

4. Average revenue per item

```
SELECT AVG([Items Cost] / [Items Sold]) AS [Avg Revenue per Item]
FROM Sales;
```

5. Average revenue per quarter

```
SELECT AVG([Revenue]) AS [Avg Revenue per Quarter]
FROM
(SELECT Quarter, SUM([Items Cost]) AS Revenue FROM Sales GROUP BY Quarter);
```

7. Obtained Results

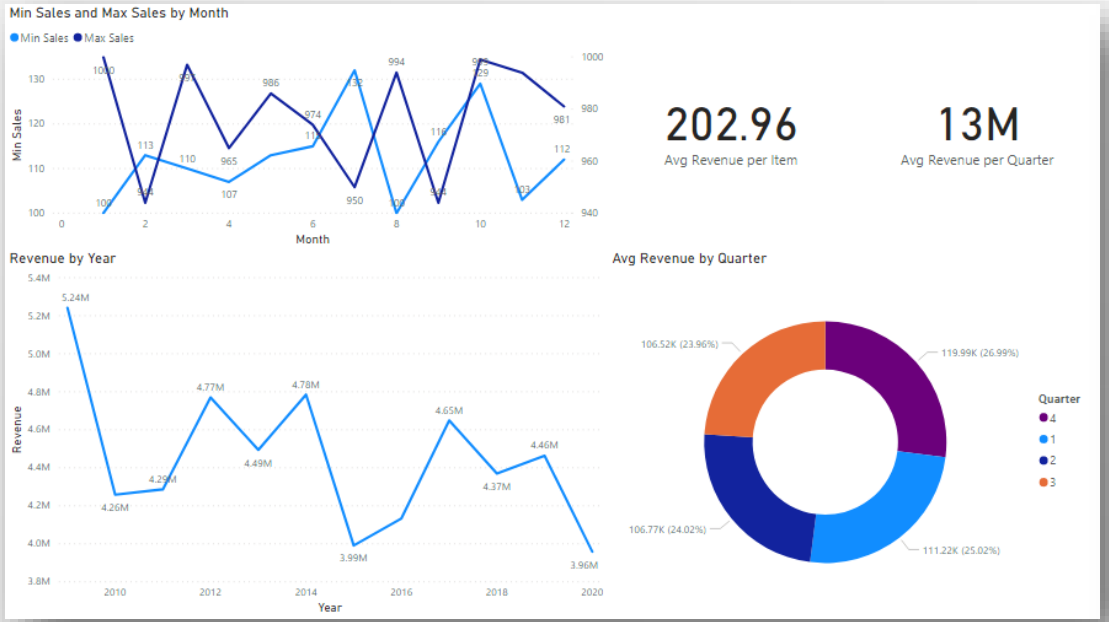
Data Subset DS	t_1	4	5	6	4	5	6	4	5	6	4	5	6
	t_2	10	10	10	11	11	11	12	12	12	15	15	15
1	$\|DS_1\| = 4$	P	P	P	P	P	P	P	P	P	P	P	P
2	$\|DS_2\| = 12$	L	L	L	L	L	L	B	B	B	B	B	B
3	$\|DS_3\| = 481$	L	L	L	L	L	L	L	L	L	L	L	L
4	$\|DS_4\| = 1$	C	C	C	C	C	C	C	C	C	C	C	C
5	$\|DS_4\| = 1$	C	C	C	C	C	C	C	C	C	C	C	C

- B stands for a bar chart
- C stands for a card
- L stands for a line chart
- P stands for a pie chart

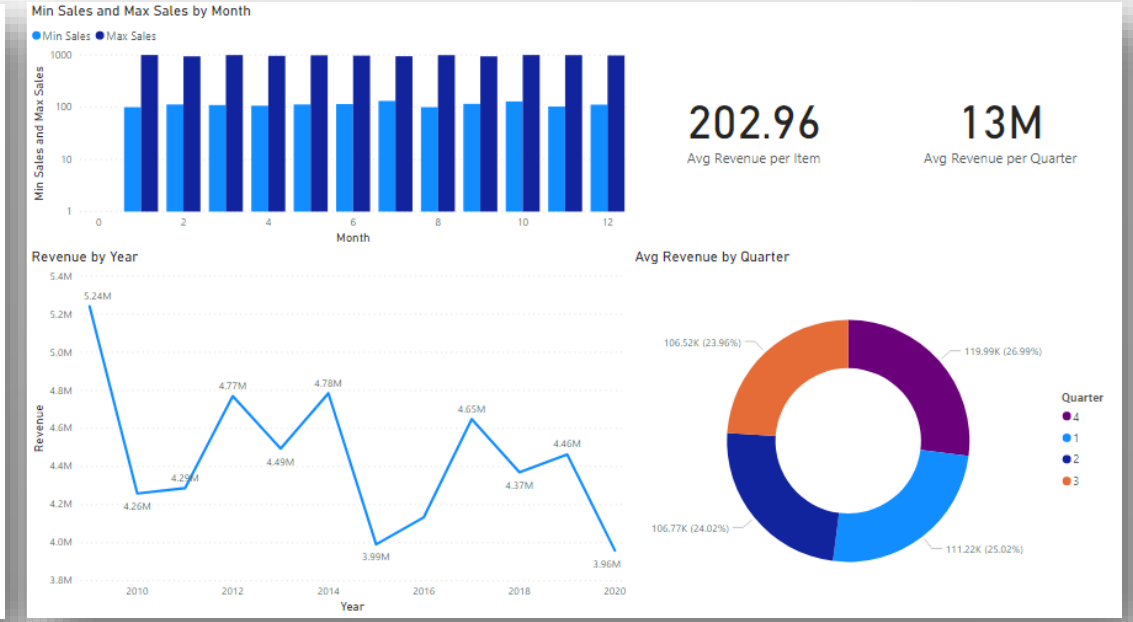
PLLCC	**PBLCC**
DS1 – Pie Chart DS2 – Line Chart DS3 – Line Chart DS4 – Card DS5 – Card	DS1 – Pie Chart DS2 – Bar Chart DS3 – Line Chart DS4 – Card DS5 – Card

7. Obtained Results

PLLCC



PBLCC



Conclusion and Future Work

1. The dashboard design problem is discussed
2. The dashboard design approach is proposed
3. Transformation from the star schema into flat data structure is described
4. The function, which maps data subsets to visualizations and helps to elaborate recommendations for dashboard design, is proposed
5. Sample dataset, data subsets based on SQL queries, and proposed visual charts are considered
6. Obtained dashboards created in Microsoft Power BI are demonstrated
7. Future research includes software implementation of the proposed approach, as well as study of visualizations relative sizes and issues of their placement on a limited dashboard space

THANK YOU FOR YOUR ATTENTION!