Data Analysis

Compiled by Jonathan Ogwumike
HiiT Plc 2023

Definition

WHAT IS DATA ?

Sounds — Images — Segments — DATA — Words — Numbers
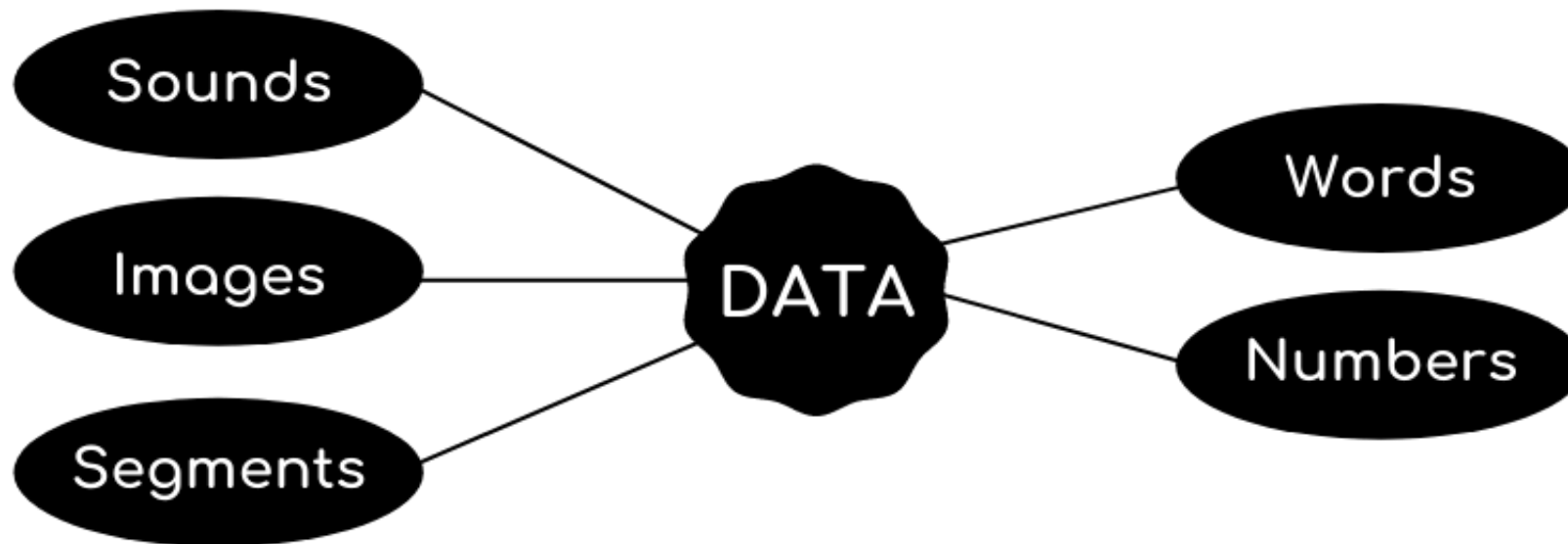
Definition

- ❑ Data is a collection of raw facts representing things or events that have happened.

Definition

❑ Data analysis is defined as a process of cleaning, transforming, and modeling data to discover useful information for business decision-making.

❑ The purpose of Data Analysis is to extract useful information from data and taking decisions based upon the data analysis result.

❑ It involves the use of analytical tools to derive insights from raw data.

DS∘DA∘BI

❑ Data science combines multiple fields, including statistics, scientific methods, artificial intelligence (AI), and data analysis, to extract value from data. Those who practice data science are called data scientists, and they combine a range of skills to analyze data collected from the web, smartphones, customers, sensors, and other sources to derive actionable insights.

DS∘DA∘BI

❑Data Analysis is the use of analytical tools to derive insights from raw data, it is a branch of DS.

❑Business Intelligence (BI) is the use of computing technologies, applications, and practices for the collection, integration, analysis, and presentation of business information. Business Intelligence solutions provide current, historical, and predictive views of internally structured data for products and departments by establishing more effective decision-making and strategic operational insights.

DA Methods

Text Analysis

Statistical Analysis

Diagnostic Analysis

Predictive Analysis

Prescriptive Analysis

Cluster analysis.

Cohort analysis.

Regression analysis.

Factor analysis.

Neural Networks.

Data Mining.

Importance of DA

❏ Managers with inaccurate or incomplete information will tend, on average, to make worse decisions than if they had better information.

❏ Creators of financial models recognize this as "garbage in, garbage out." BI attempts to solve this problem by analyzing current data that is ideally presented on a dashboard of quick metrics designed to support better decisions.

Data science encompasses preparing data for analysis, including cleansing, aggregating, and manipulating the data to perform advanced data analysis. Analytic applications and data scientists can then review the results to uncover patterns and enable business leaders to draw informed insights.

Data Analysis Process

The **Data Analysis Process** is nothing but gathering information by using a proper application or tool which allows you to explore the data and find a pattern in it. Based on that information and data, you can make decisions, or you can get to make conclusions.

Data Analysis consists of the following phases:

Data Requirement Gathering

Data Collection

Data Cleaning

Data Analysis

Data Interpretation

Data Visualization

Why Analyze your data?

❑Power BI allows you to view, analyze and visualize huge quantities of Data.

❑Power BI gives you the flexibility to handle large files which cannot be otherwise done with Excel.

❑Power BI uses powerful compression algorithms to import and cache the data within the PBIX file.

❑It is easy to create stunning visuals with Power BI. It makes things visually appealing.

❑It has easy drag and drop functionality, with features that allow you to easily copy all formatting across similar visualizations.

❑Power BI has exceptional Excel integration, allowing easy sharing of files, queries, data models and reports.

Data Analysis Tools

Data analysis tools make it easier for users to process and manipulate data, analyze the relationships and correlations between data sets, and it also helps to identify patterns and trends for interpretation.

Some of the most popular notebooks are Jupyter, RStudio, and Zeppelin. Notebooks are very useful for conducting analysis, but have their limitations when data scientists need to work as a team. Data science platforms were built to solve this problem.

Data Analysis Tools

R

- ❏ R Prep
- ❏ R Markdown
- ❏ R Studio

SQL

- ❏ SQL Server
- ❏ MySQL
- ❏ Azure

Python

- ❏ Anaconda
- ❏ Jupyter notebook
- ❏ ipython

Microsoft Excel

- ❏ Query Editor
- ❏ Get and transform
- ❏ Analysis toolpack

Power BI

- ❏ Power BI Online
- ❏ Power BI Desktop

Tableau

- ❏ Tableau Desktop
- ❏ Tableau Online
- ❏ Tableau Public (Free)
- ❏ Tableau Server

Summary:

Data analysis means a process of cleaning, transforming and modeling data to discover useful information for business decision-making

Types of Data Analysis are Text, Statistical, Diagnostic, Predictive, Prescriptive Analysis

Data Analysis consists of Data Requirement Gathering, Data Collection, Data Cleaning, Data Analysis, Data Interpretation, Data Visualization

Summary:

The six phases of the data analysis process:

ask,

prepare,

process,

analyze,

share,

and act.

Summary:

Data Life Cycle:

Plan

Capture

Manage

Analyze

Archive

Destroy

Key data analyst tools

As you are learning, the most common programs and solutions used by data analysts include spreadsheets, query languages, and visualization tools. In this reading, you will learn more about each one. You will cover when to use them, and why they are so important in data analytics.

Spreadsheets

Data analysts rely on spreadsheets to collect and organize data. Two popular spreadsheet applications you will probably use a lot in your future role as a data analyst are Microsoft Excel and Google Sheets.

Spreadsheets structure data in a meaningful way by letting you

•Collect, store, organize, and sort information

•Identify patterns and piece the data together in a way that works for each specific data project

•Create excellent data visualizations, like graphs and charts.

•Spreadsheets enable data analysts to store, organize, sort, and filter data. This helps them see patterns, group information, and easily find the information they need.

Databases and query languages

A database is a collection of structured data stored in a computer system. Some popular Structured Query Language (SQL) programs include MySQL, Microsoft SQL Server, and BigQuery.

Query languages

•Allow analysts to isolate specific information from a database(s)

•Make it easier for you to learn and understand the requests made to databases

•Allow analysts to select, create, add, or download data from a database for analysis

Visualization tools

Data analysts use a number of visualization tools, like graphs, maps, tables, charts, and more.
Three popular visualization tools are Power BI, Tableau, and Looker Studio.
These tools
•Turn complex numbers into a story that people can understand
•Help stakeholders come up with conclusions that lead to informed decisions and effective
business strategies
•Have multiple features
- **Tableau**'s simple drag-and-drop feature lets users create interactive graphs in dashboards
and
worksheets
- **Looker** communicates directly with a database, allowing you to connect your data right to the
visual
tool you choose

Programming Languages tools

A career as a data analyst also involves using programming languages, like R and Python, which are used a lot for statistical analysis, visualization, and other data analysis.

Key takeaway
You have a lot of tools as a data analyst. This is a first glance at the possibilities, and you will explore many of these tools in-depth throughout this program.

Data Visualization

- There's a reason why you want to visualize your data.

  You want to tell a story.

  What is that story?

- What do you want your data to tell people?

- *Example*

- Just imagine you are the sales manager of a thriving business and you want to share the success of a recent marketing campaign.

  That's your story.

- By understanding what you want your data to say, you can understand how you should visualize it.

Importance of visualization

Data visualization is an important part of any data analysis project.

Your reports should be able to communicate key insights from your analysis to any audience.

There are many different scenarios where large amounts of data must be displayed to an audience – a business may need to present sales figures to their directors, a research team may need to display their findings to investors, or a teacher may need to display statistics to their students for example.

Data visualization allows this information to be displayed in an easy to read format that is both attractive and functional.

Data Visualization - Types

Popular types of data visualization include graphs, charts, and infographics.

Power BI gives you access to loads of visuals out of the box and more options via 3rd party add-ons
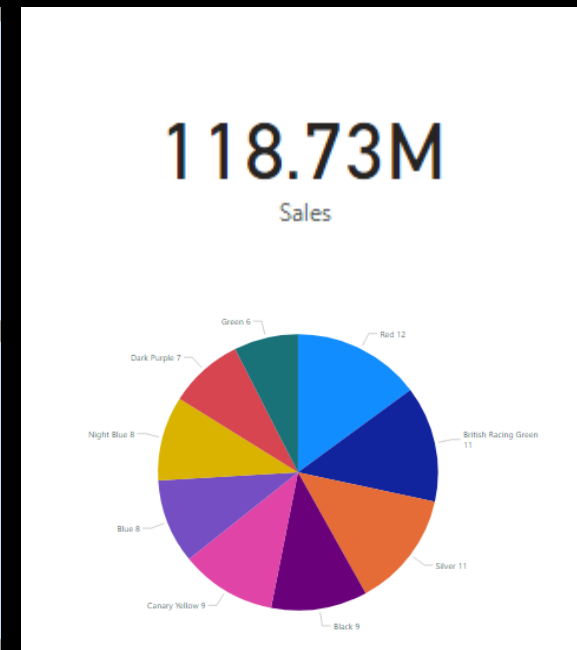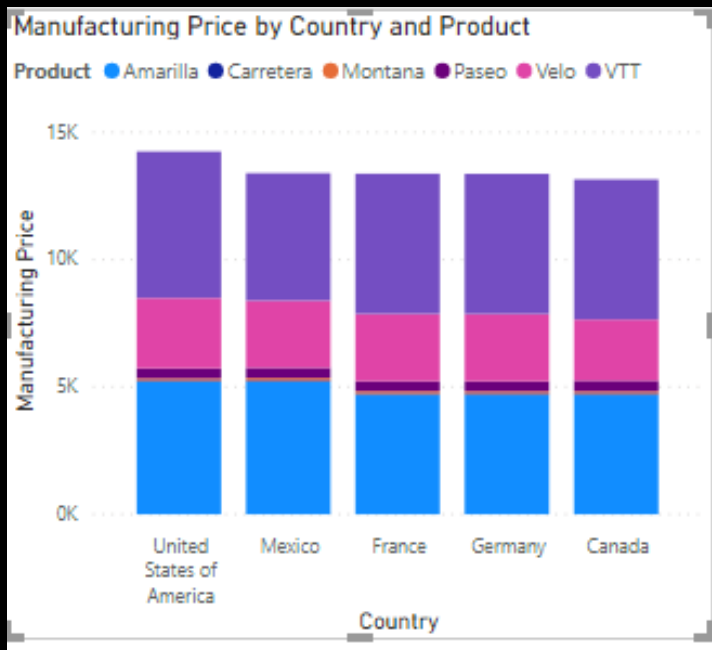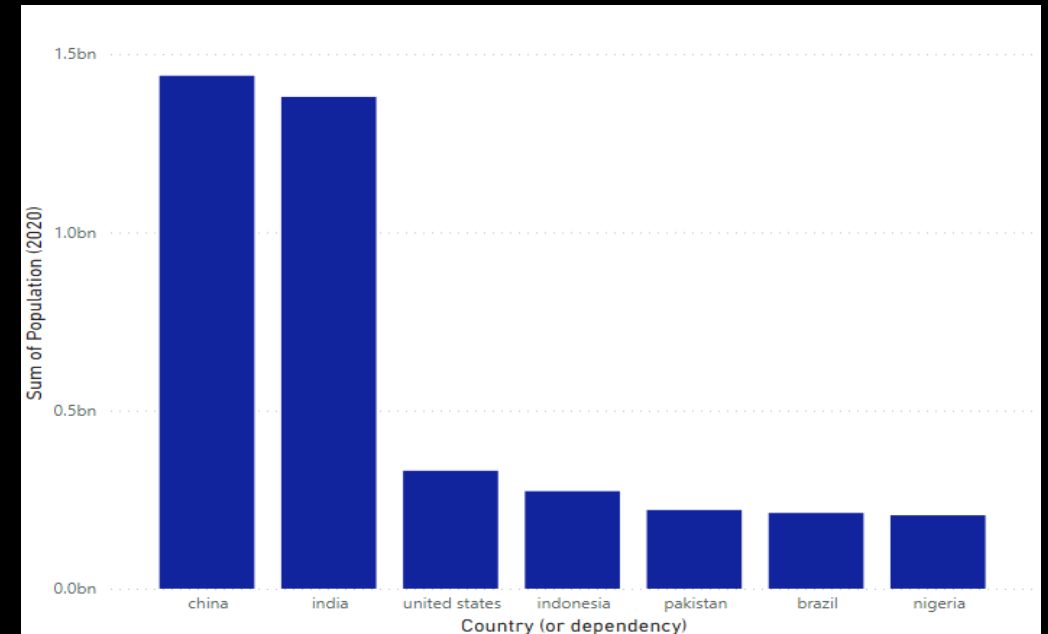
DATA VISUALIZATION
CHOOSING THE RIGHT CHART

COMPARE DIMENSIONS OR CATEGORIES

Bar charts compare metrics using vertical or horizontal bars, and allow you to compare discrete quantities of different categories.

Tip: To compare differences between categories, use bar charts only with a reasonable number of data points, such as up to 12, to keep comparisons visually clear.
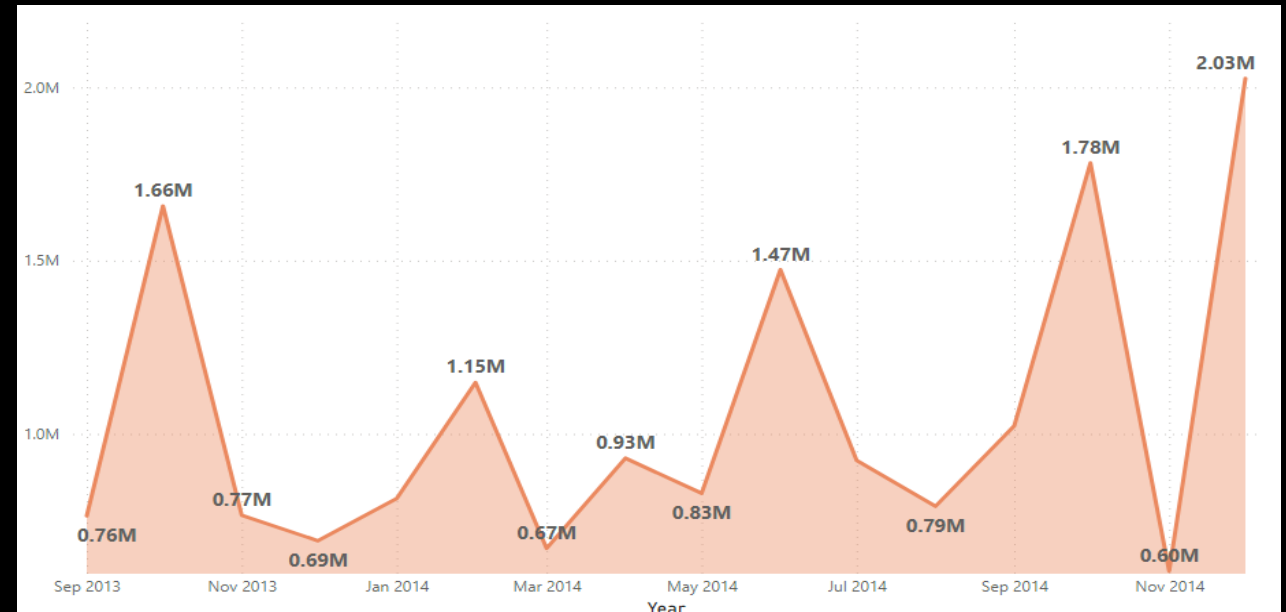
DATA VISUALIZATION
CHOOSING THE RIGHT CHART

A **line chart** can easily help depict the relationship between two or more variables over time.

Similarly, an area chart can help you discover correlations within a dimension.
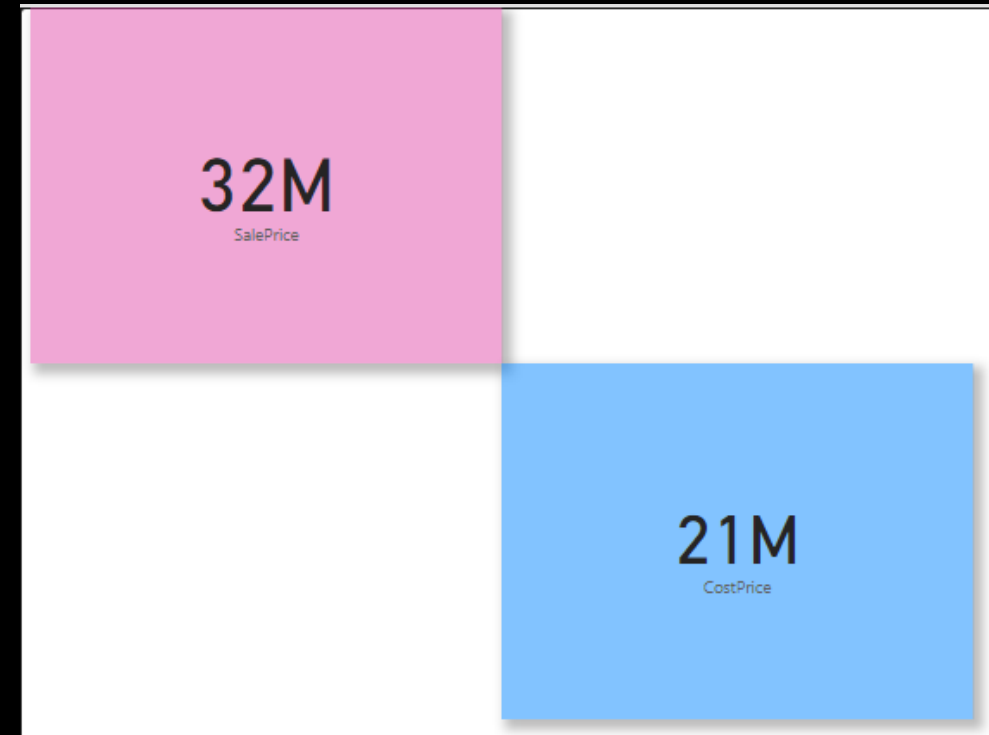
DATA VISUALIZATION
CHOOSING THE RIGHT CHART

HIGHLIGHT A SINGLE IMPORTANT METRIC

A **card** shows a total for a single metric.

This won't be grouped with any dimension, but will be impacted by filter controls.

Cards may include a comparison to a previous time period, indicating increases or decreases.

32M
SalePrice

21M
CostPrice

VIEW THE COMPOSITION OF A SET OF DATA

Pie charts compare the parts of a whole, and are most helpful when comparing only two or three values.

Use cautiously; be careful not to present a comparison of two pie charts without indicating that the size of the pie may be different.
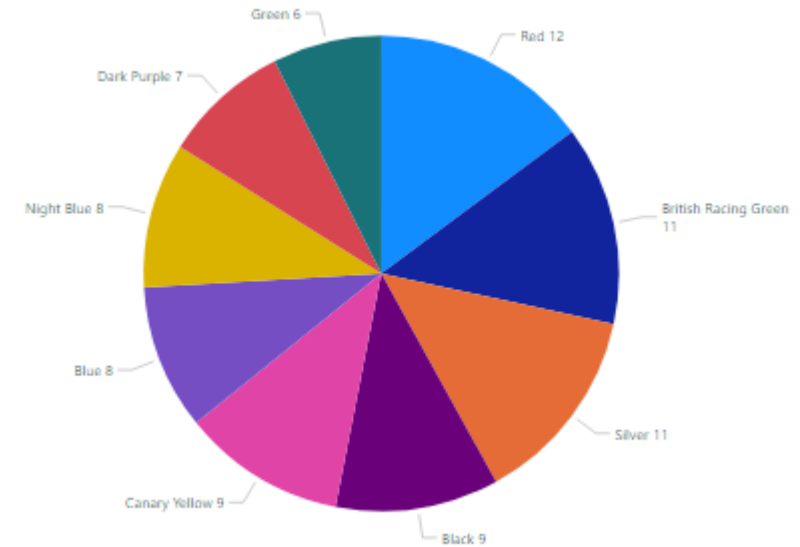
DATA VISUALIZATION
CHOOSING THE RIGHT CHART

VISUALIZE DATA IN THE REAL WORLD

A **map** provides an easy way to visualize how a measurement varies across a geographic area. To use a map, your data set needs to have LOCATION data, such as city, latitude, longitude, or country.
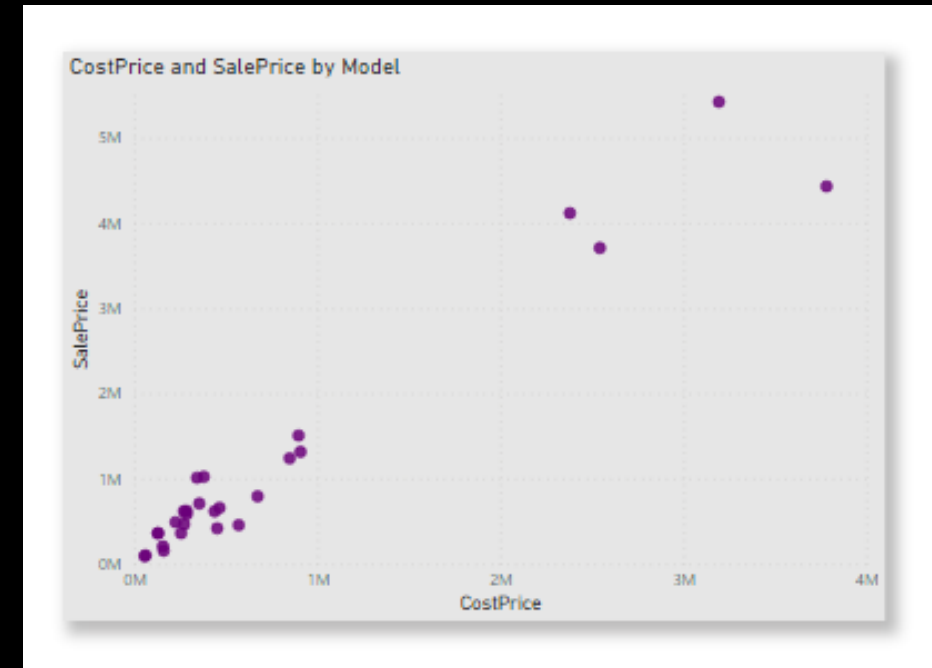
DATA VISUALIZATION
CHOOSING THE RIGHT CHART

Scatter and bubble charts help you identify a correlation or lack of correlation between data points.

These charts show your data as points or circles on a graph using X (left to right) and Y (top to bottom) axes.

Scatter charts can include a trend line that shows how the variables in the chart are related.
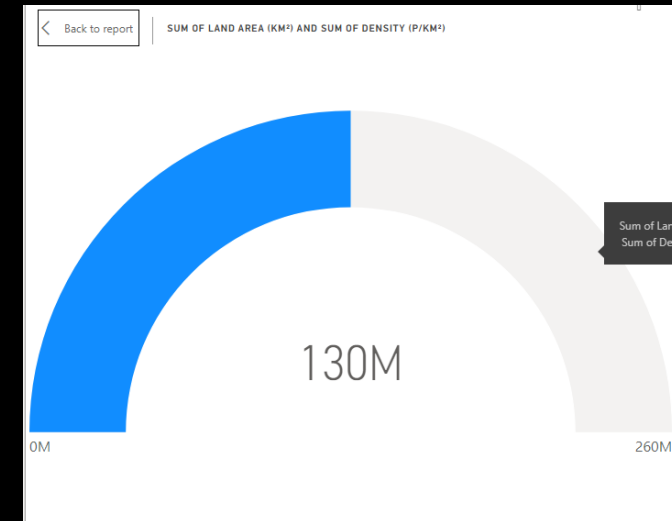

CostPrice and SalePrice by Model

DATA VISUALIZATION
CHOOSING THE RIGHT CHART

A **GUAGE chart** gives you a way to quickly see how well a given metric is performing against target benchmarks.

KPI charts are often used in dashboards to provide meter or gauge-like widgets that monitor various "health" or performance KPI (key performance indicators).

DATA VISUALIZATION
CHOOSING THE RIGHT CHART

1. The viewer wants to see how revenue has changed over time?

2. Viewers want to quickly compare sales of several different product categories.

3. Viewers want to identify the cities where most website visitors live.