Data Era

What Happens in 60 Seconds

Data is new oil

David Parkins
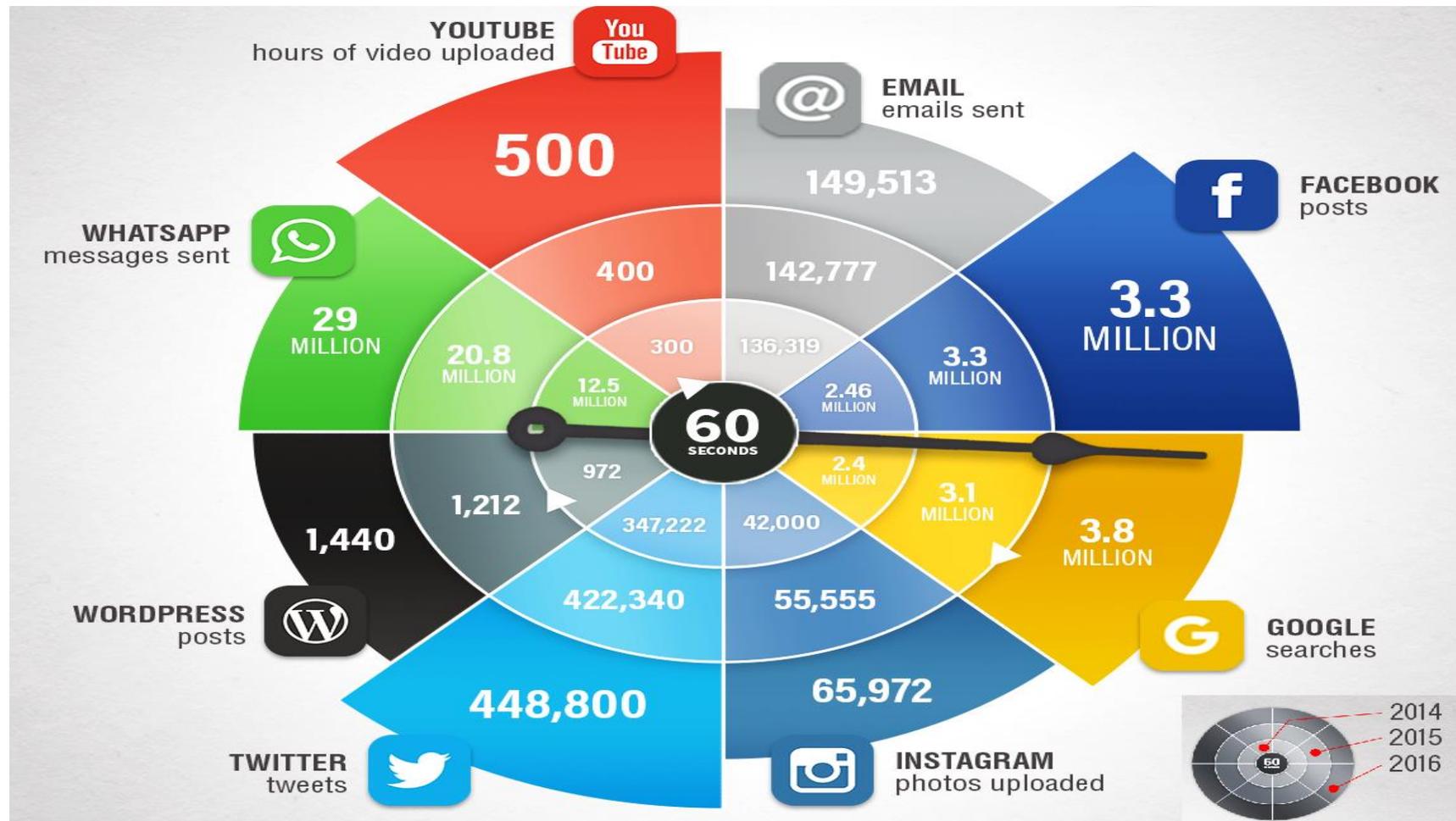
Data Sources

Text

Images

Videos

Clickstreams

User conversations

Social media platform

IOT

legacy database

Data from Dirty to Clean

#	Id	Name	Birthday	Gender	IsTeacher?	#Students	Country	City
1	111	John	31/12/1990	M	0	0	Ireland	Dublin
2	222	Mery	15/10/1978	F	1	15	Iceland	
3	333	Alice	19/04/2000	F	0	0	Spain	Madrid
4	444	Mark	01/11/1997	M	0	0	France	Paris
5	555	Alex	15/03/2000	A	1	23	Germany	Berlin
6	555	Peter	1983-12-01	M	1	10	Italy	Rome
7	777	Calvin	05/05/1995	M	0	0	Italy	Italy
8	888	Roxane	03/08/1948	F	0	0	Portugal	Lisbon
9	999	Anne	05/09/1992	F	0	5	Switzerland	Geneva
10	101010	Paul	14/11/1992	M	1	26	Ytali	Rome

← Missing values

← Invalid values

← Misfielded values

← Misspellings

Uniqueness Formats Attribute dependencies

Dirty

Name Hussein Hakeem Address Number 22 Fioye Crescent Surulere Lagos Age 17 Gender Male
Name Arojoye Samuel Address 11 Omolade Close Omole Estate Lagos Age 16 Gender Male
Name Alex Ezurum Address 1 Adamu Lane, Abuja Age 14 Gender Male
Name Susan Nwaimo Address Number 58 Yaba Street, Kaduna State Age 16 Gender Female
Name Ajao Opeyemi Address No12 Olubunmi Street, Abeokuta Age 18 Gender Female
Name Banjoko Adebusola Address 34 Ngige Street, Ugheli, Delta Age 14 Gender Female
Name Muhammed Olabisi Address 13, ICAN road, Enugu Age 12 Gender Female
Name Oluwagbemi Mojisola Address ACCA Lane, Onitsha Age 13 Gender Female

Clean

Name	Address	Age	Gender
Hussein Hakeem	Number 22 Fioye Crescent Surulere Lagos	17	Male
Arojoye Samuel	11 Omolade Close Omole Estate Lagos	16	Male
Alex Ezurum	1 Adamu Lane, Abuja	14	Male
Susan Nwaimo	Number 58 Yaba Street, Kaduna State	16	Female
Ajao Opeyemi	No12 Olubunmi Street, Abeokuta	18	Female
Banjoko Adebusola	34 Ngige Street, Ugheli, Delta	14	Female
Muhammed Olabisi	13, ICAN road, Enugu	12	Female
Oluwagbemi Mojisola	ACCA Lane, Onitsha	13	Female

END-USER
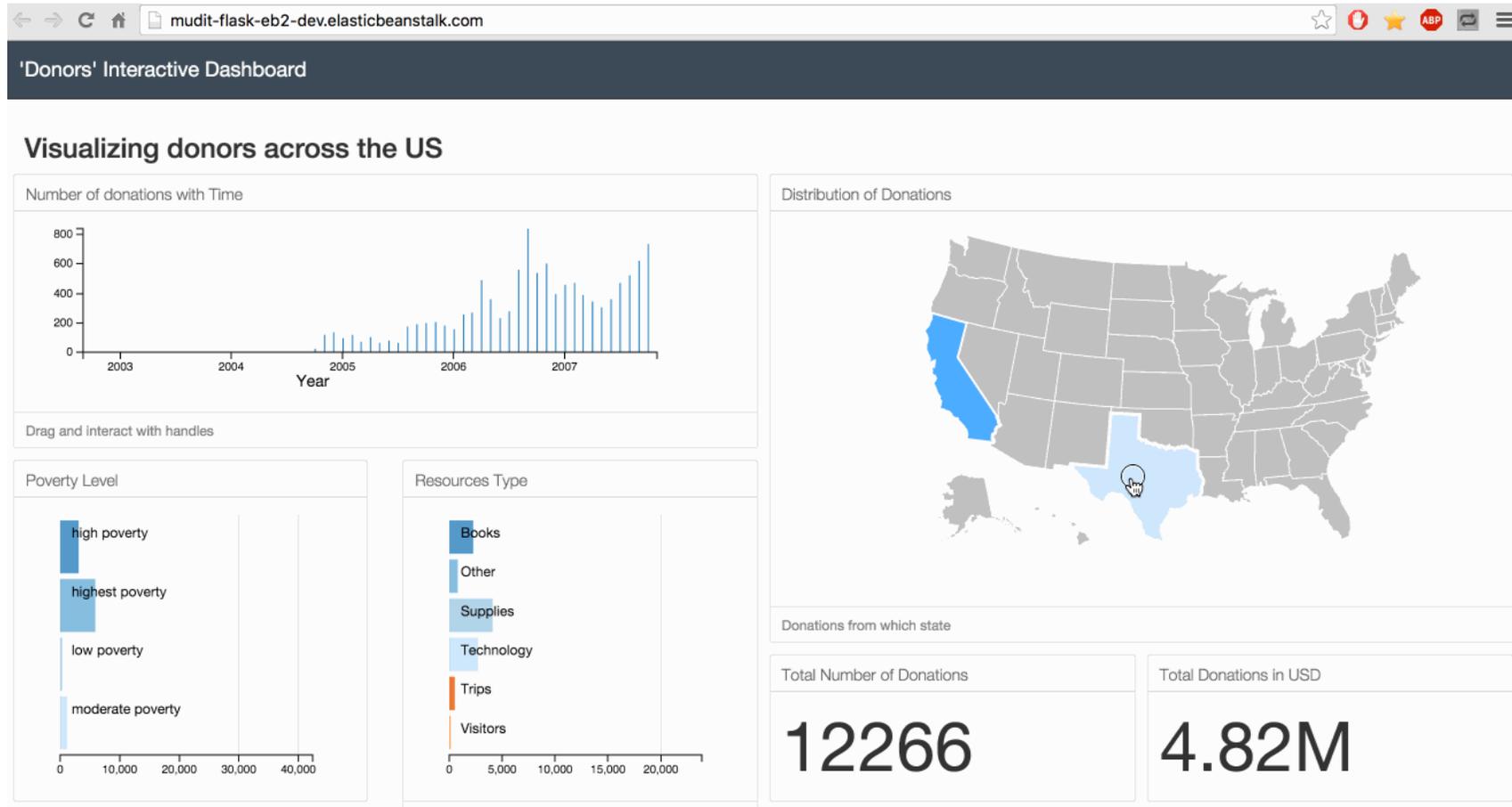
Data Professional Team
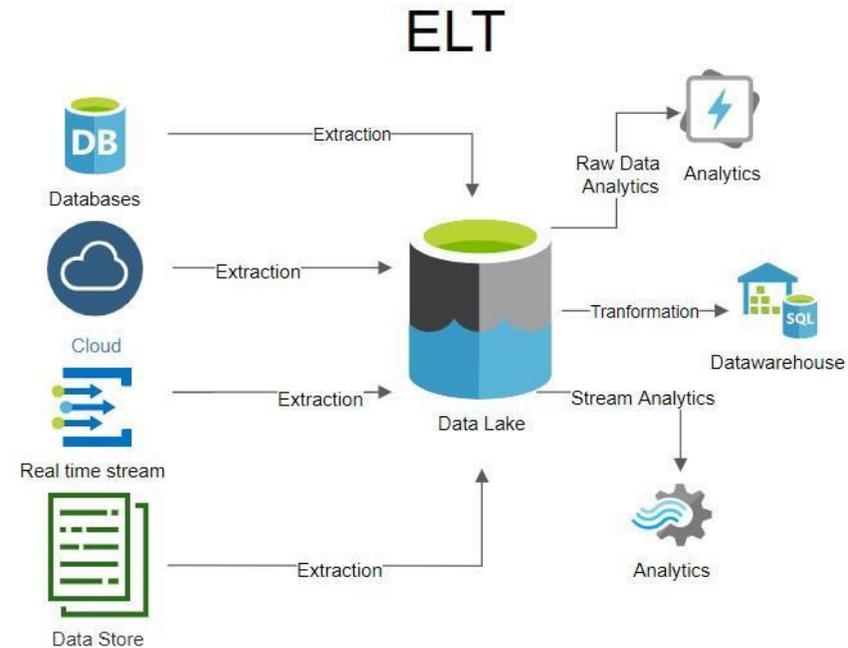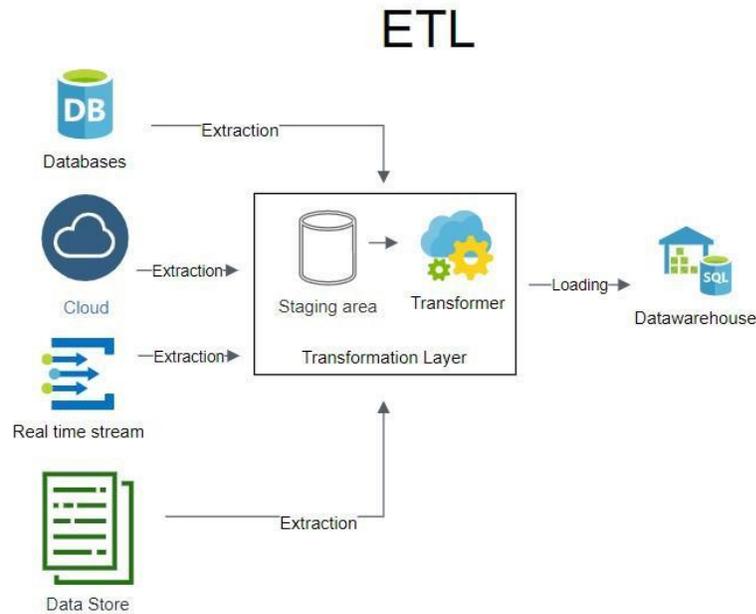
Data Engineers

Data Analysts

Data Scientists

Business Intelligence

Data Engineering

- Extract, integrate, and organize data from disparate sources
- Clean, transform, and prepare data
- Design, store, and manage data in data repositories

Data Engineers Skills

- Good knowledge of programming
- Sound knowledge of systems and technology architectures
- In-depth understanding of relational databases and non-relational data stores

Data Analysts

- Inspect and clean data for deriving insights

- Identify correlations, find patterns, and apply statistical methods to analyze and mine data

- Visualize data to interpret and present the findings of data analysis
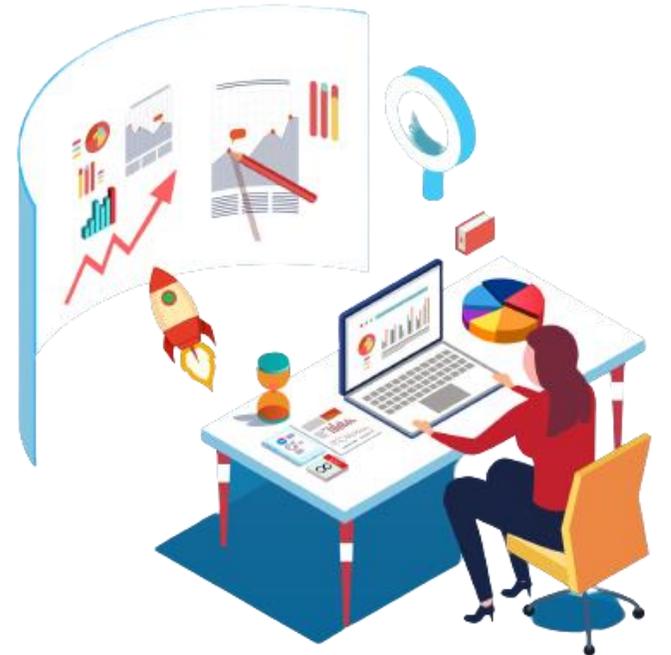
Data Analysts Skills

- Good knowledge of spreadsheets, writing queries, and using statistical tools to create charts and dashboards

- Programming skills

- Strong analytical and story-telling skills

Data Science

- Analyze data for actionable insights
- Create predictive models using Machine Learning and Deep Learning

Data Scientists Skills

- Knowledge of Mathematics and Statistics

- Understanding of programming languages, databases, and building data models

- Domain knowledge

Business Intelligence

Business Analysts leverage the work of Data Analysts and Data Scientists to look at possible implications for their business and the actions they need to take or recommend.

Business Intelligence Skills

- Focus on market forces and external influences that shape their business

- Organize and monitor data on different business functions

- Explore data to extract insights and actionable that improve business performance

To Summarize

Data Engineering converts raw data into usable data

Data Analytics use this data to generate insights

Data Scientists use Data Analytics and Data Engineering to predict the future using data from the past

Business Analysts and Business Intelligence Analysts use these insights and predictions to drive decisions that benefit and grow their business

Spreadsheet Applications

- Most commonly used

- Part of the Office suite

- Web-based version called Excel for the web

Spreadsheet Applications

- Free with a Google account

- Web-based application

- Integrates with other Google apps

Spreadsheet Applications

- Free and open-source desktop spreadsheet application

Spreadsheet Applications

Microsoft Excel

Google Sheets

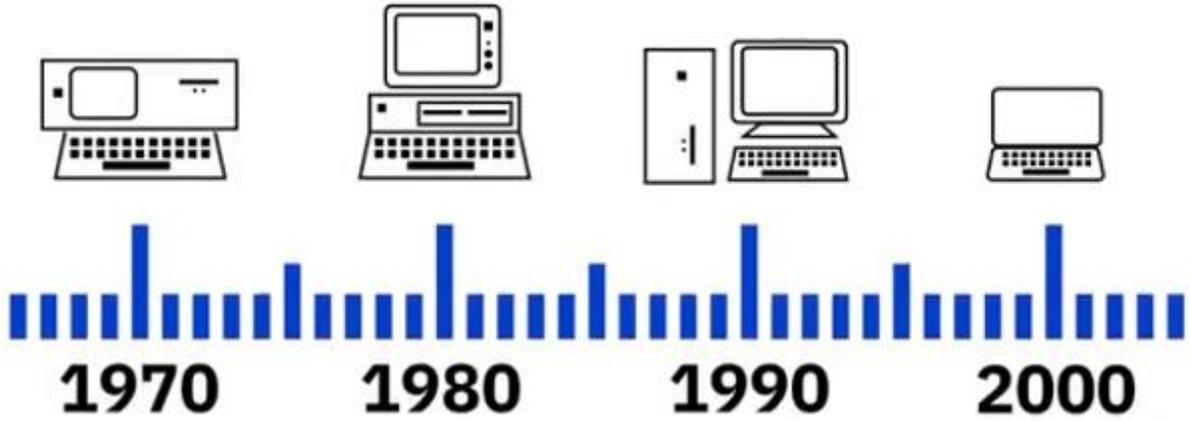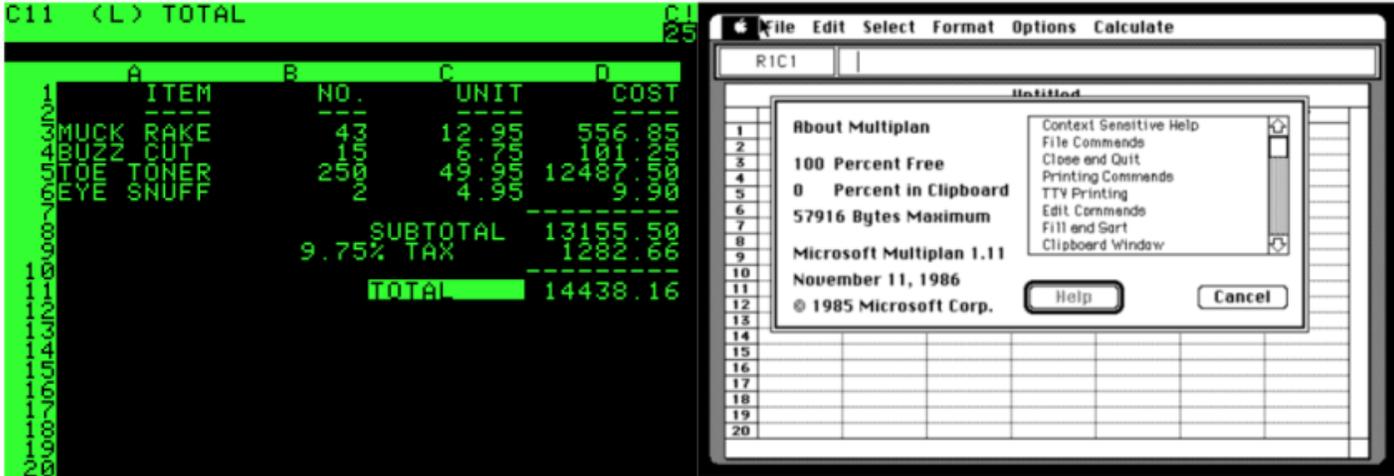Libre Office Calc
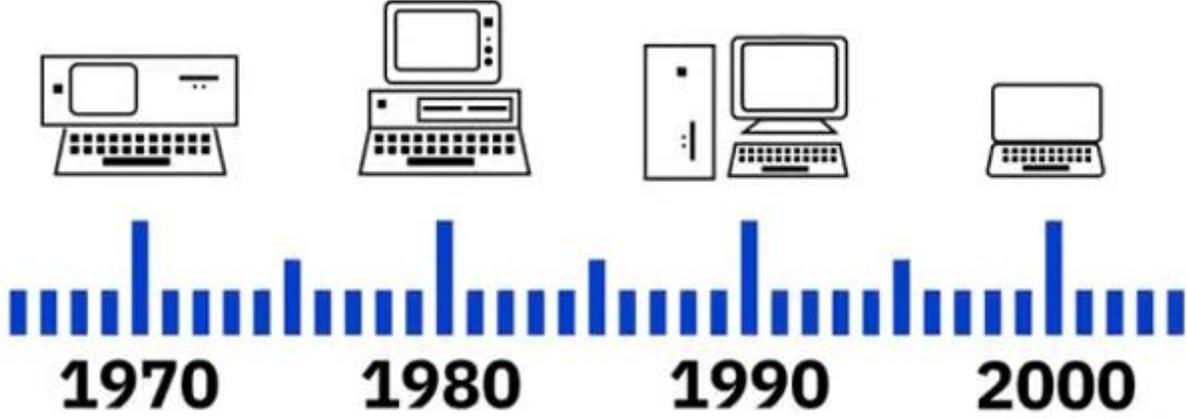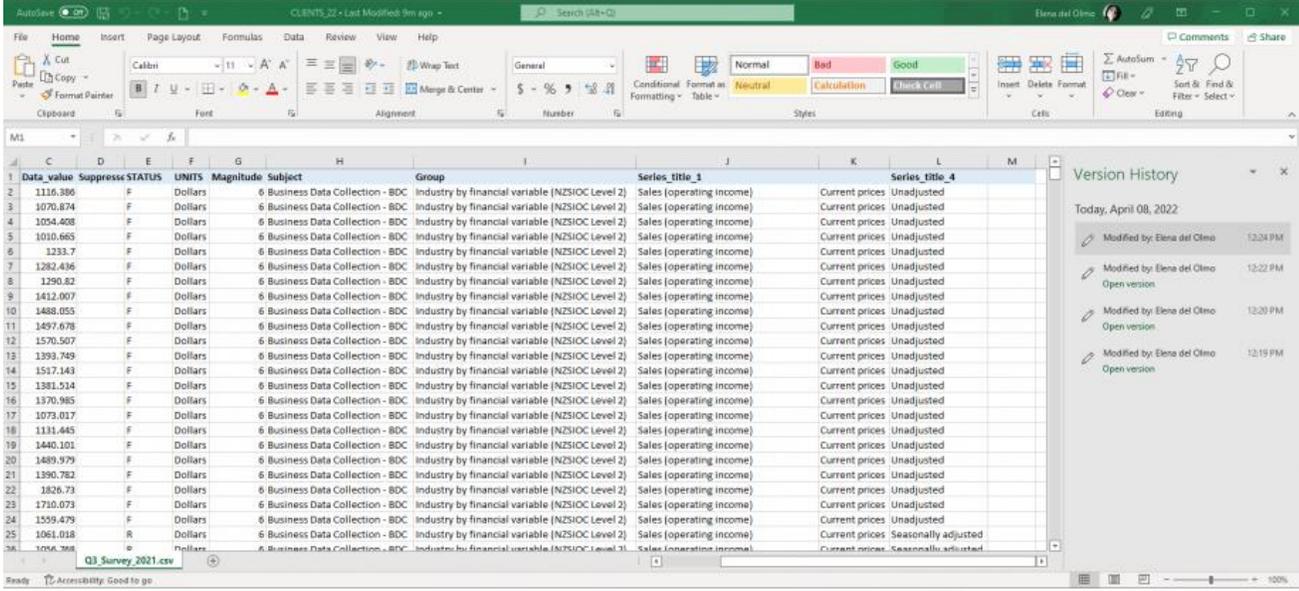
Zoho Sheet

Open Office Calc

Quip

Smartsheet

Apple Numbers

Advantages of using Spreadsheets

- Accurate calculations
- Automatic calculations
- Organize and access data
- Format, filter, and sort data
- Edit, Undo, and Error-check
- Analyze data
- Create charts, graphs, and reports

# Advantages of using Spreadsheets

Advantages of using Spreadsheets

The most common business uses for spreadsheet applications include the following:

- Data Entry and Storage
- Comparing Large Datasets
- Modelling and Planning
- Charting
- Identifying Trends
- Flowcharts for Business Processes
- Tracking Business Sales

- Financial Forecasting
- Statistical Analysis
- Profit and Loss Accounting
- Budgeting
- Forensic Auditing
- Payroll and Tax Reporting
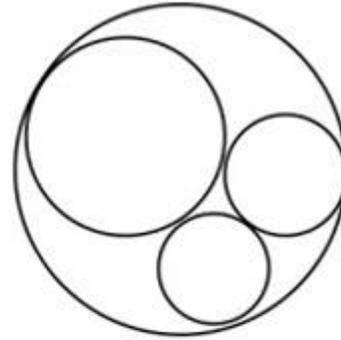- Invoicing
- Scheduling

Non-business uses for spreadsheet applications include the following:

- Personal Expenses
- Household Budgeting
- Recipe Library
- Fitness Tracking
- Calorie Counting & Weight Monitoring

- Sports Leagues such as Fantasy Football
- Cataloging Music Libraries
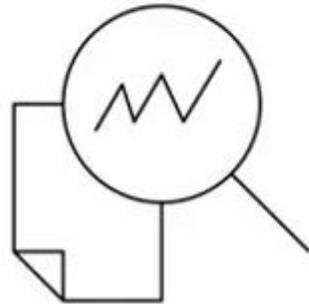- Contact Lists
- Shopping Lists
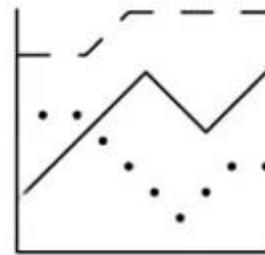
How a Data Analyst Uses Spreadsheets
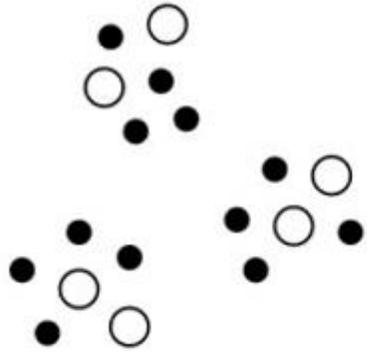
Collecting Data

Cleaning Data

Analyzing Data

Visualizing Data