Google data analytics professional course

Week-1

Case Study: New data perspectives

Ask

- *What do you think new employees need to learn to be successful in their first year on the job?*
- *Have you gathered data from new employees before? If so, may we have access to the historical data?*
- *Do you believe managers with higher retention rates offer new employees something extra or unique?*
- *What do you suspect is a leading cause of dissatisfaction among new employees?*
- *By what percentage would you like employee retention to increase in the next fiscal year?*

Prepare

*It all started with solid **preparation**. The group built a timeline of three months and decided how they wanted to relay their progress to interested parties. Also during this step, the analysts identified what data they needed to achieve the successful result they identified in the previous step - in this case, the analysts*

chose to gather the data from an online survey of new employees. These were the things they did to prepare:

- They developed specific questions to ask about employee satisfaction with different business processes, such as hiring and onboarding, and their overall compensation.
- They established rules for who would have access to the data collected - in this case, anyone outside the group wouldn't have access to the raw data, but could view summarized or aggregated data. For example, an individual's compensation wouldn't be available, but salary ranges for groups of individuals would be viewable.
- They finalized what specific information would be gathered, and how best to present the data visually. The analysts brainstormed possible project- and data-related issues and how to avoid them.
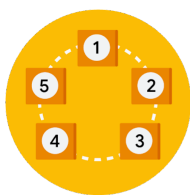
Process

The group sent the survey out. Great analysts know how to respect both their data and the people who provide it. Since employees provided the data, it was important to make sure all employees gave their consent to participate. The data analysts also made sure employees understood how their data would be **collected, stored, managed, and protected**. Collecting and using data ethically is one of the responsibilities of data analysts. In order to maintain confidentiality and protect and store the data effectively, these were the steps they took:

- They restricted access to the data to a limited number of analysts.
- They cleaned the data to make sure it was complete, correct, and relevant. Certain data was aggregated and summarized without revealing individual responses.

- They uploaded raw data to an internal data warehouse for an additional layer of security.

Analyze

Then, the analysts did what they do best: analyze! From the completed surveys, the data analysts **discovered** that an employee's experience with certain processes was a key indicator of overall job satisfaction. These were their findings:

- Employees who experienced a long and complicated hiring process were most likely to leave the company.
- Employees who experienced an efficient and transparent evaluation and feedback process were most likely to remain with the company.

The group knew it was important to **document** exactly what they found in the analysis, no matter what the results. To do otherwise would diminish trust in the survey process and reduce their ability to collect truthful data from employees in the future.

Share

Just as they made sure the data was carefully protected, the analysts were also careful **sharing the report**. This is how they shared their findings:

- They shared the report with managers who met or exceeded the minimum number of direct reports with submitted responses to the survey.
- They presented the results to the managers to make sure they had the full picture.
- They asked the managers to personally deliver the results to their teams.

This process gave managers an opportunity to **communicate the results** with the right context. As a result, they could have productive team conversations about next steps to improve employee engagement.

Act

The last stage of the process for the team of analysts was to work with leaders within their company and decide how best to **implement changes and take actions** based on the findings. These were their recommendations:

- Standardize the hiring and evaluation process for employees based on the most efficient and transparent practices.
- Conduct the same survey annually and compare results with those from the previous year.

A year later, the same survey was distributed to employees. Analysts anticipated that a comparison between the two sets of results would indicate that the action plan worked. Turns out, the changes improved the retention rate for new employees and the actions taken by leaders were successful!

- https://online.hbs.edu/blog/post/business-analytics-examples

Learning Log: Consider how data analysts approach tasks

The six phases of the data analysis process:

1. **Ask** questions and define the problem.
2. **Prepare** data by collecting and storing the information.
3. **Process** data by cleaning and checking the information.
4. **Analyze** data to find patterns, relationships, and trends.
5. **Share** data with your audience.
6. **Act** on the data and use the analysis results.

★ The analysts asked questions to define both the issue to be solved and what would equal a successful result.

★ Next, they prepared by building a timeline and collecting data with employee surveys that were designed to be inclusive.

★ They processed the data by cleaning it to make sure it was complete, correct, relevant, and free of errors and outliers.

★ They analyzed the clean employee survey data.

★ Then the analysts shared their findings and recommendations with team leaders.

★ Afterward, leadership acted on the results and focused on improving key areas.

Data and gut instinct

Detectives and data analysts have a lot in common. Both depend on facts and clues to make decisions. Both collect and look at the evidence. Both talk to

people who know part of the story. And both might even follow some footprints to see where they lead. Whether you're a detective or a data analyst, your job is all about following steps to collect and understand facts.

Gut instinct is an intuitive understanding of something with little or no explanation.

Data + business knowledge = mystery solved

In addition, try asking yourself these questions about a project to help find the perfect balance:

- What kind of results are needed?
- Who will be informed?
- Am I answering the question being asked?
- How quickly does a decision need to be made?

Ecosystem

- An ecosystem is a group of elements that interact with one another.
- So when you think about data, data analysis and the data ecosystem, it's important to understand that all of these things fit under the data analytics umbrella.

Data alone will never be as powerful as data combined with human experience, observation, and sometimes even intuition.

Big data analytics life cycle

Authors Thomas Erl, Wajid Khattak, and Paul Buhler proposed a big data analytics life cycle in their book, **Big Data Fundamentals: Concepts, Drivers & Techniques**. Their life cycle suggests phases divided into nine steps:

1. Business case evaluation

2. Data identification
3. Data acquisition and filtering
4. Data extraction
5. Data validation and cleaning
6. Data aggregation and representation
7. Data analysis
8. Data visualization
9. Utilization of analysis results

Week-2

About analytical thinking

Learning Log: Reflect on your skills and expectations

The table has a row for each essential aspect of **analytical skills**:

- **Curiosity:** a desire to know more about something, asking the right questions
- **Understanding context:** understanding where information fits into the "big picture"
- **Having a technical mindset:** breaking big things into smaller steps
- **Data design:** thinking about how to organize data and information
- **Data strategy:** thinking about the people, processes, and tools used in data analysis

Week-3

Follow the data life cycle

Variations of the data life cycle

1. **Plan:** Decide what kind of data is needed, how it will be managed, and who will be responsible for it.
2. **Capture:** Collect or bring in data from a variety of different sources.
3. **Manage:** Care for and maintain the data. This includes determining how and where it is stored and the tools used to do so.
4. **Analyze:** Use the data to solve problems, make decisions, and support business goals.
5. **Archive:** Keep relevant data stored for long-term and future reference.
6. **Destroy:** Remove data from storage and delete any shared copies of the data.

How the data analysis process guides this program

Ask ○——————————— ①
- Ask effective questions
- Define the problem
- Use structured thinking
- Communicate with others

② ———————————○ **Prepare**
- Understand how data is generated and collected
- Identify and use different data formats, types, and structures
- Make sure data is unbiased and credible
- Organize and protect data

Process ○——————————— ③
- Create and transform data
- Maintain data integrity
- Test data
- Clean data
- Verify and report on cleaning results

④ ———————————○ **Analyze**
- Use tools to format and transform data
- Sort and filter data
- Identify patterns and draw conclusions
- Make predictions and recommendations
- Make data-driven decisions

Share ○——————————— ⑤
- Understand visualization
- Create effective visuals
- Bring data to life
- Use data storytelling
- Communicate to help others understand results

⑥ ———————————○ **Act**
- Apply your insights
- Solve problems
- Make decisions
- Create something new

The five key aspects to analytical thinking. They are

- *visualization,*
- *strategy,*
- *problem-orientation,*
- *correlation, and finally,*
- *big-picture and detail-oriented thinking.*

Having a **strategic** *mindset is key to staying focused and on track.*

Data analysts use a **problem- oriented** *approach in order to identify, describe, and solve problems.*

What is the root cause of a problem?

A root cause is the reason **why** *a problem occurs.*

Gap analysis *lets you examine and evaluate how a process works currently in order to get where you want to be in the future.*

> ***Warning****: Be careful not to mix up or confuse the* six stages of the data life cycle *(Plan, Capture, Manage, Analyze, Archive, and Destroy) with the six phases of the* data analysis life cycle *(Ask, Prepare, Process, Analyze, Share, and Act).*

Key data analyst tools

Spreadsheets

- *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.*

Databases and 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

- *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*

Week-4

Mastering spreadsheet basics

More spreadsheet resources-Links

- *Google Sheets Training and Help*
- *Google Sheets Cheat Sheet*
- *Microsoft Excel for Windows Training*

Structured Query Language (SQL)

SQL Guide: Getting started

What is a query?

A **query** is a request for data or information from a database. When you query databases, you use SQL to communicate your question or request. You and the database can always exchange information as long as you speak the same language.

The syntax of SQL query:

- Use **SELECT** to choose the columns you want to return.
- Use **FROM** to choose the tables where the columns you want are located.
- Use **WHERE** to filter for certain information.

Example of a query

```sql
SELECT
      first_name
FROM
      customer_data.customer_name
WHERE
      first_name = 'Tony'
```

```sql
SELECT
      customer_id,
      first_name,
      last_name
FROM
      customer_data.customer_name
WHERE
      customer_id > 0
      AND first_name = 'Tony'
      AND last_name = 'Magnolia'
```

Endless SQL possibilities

Commands

Comments are text placed between certain characters, **/*** and ***/**, or after two dashes (**--**) as shown below.

```
SELECT
      field1 /* this is the last name column */
FROM
      table -- this is the customer data table
WHERE
      field1 LIKE 'Ch%';
```

Aliases

You can also make it easier on yourself by assigning a new name or **alias** to the column or table names to make them easier to work with (and avoid the need for comments). This is done with a SQL AS clause.

```
field1 AS last_name -- Alias to make my work easier
table AS customers -- Alias to make my work easier

SELECT
      last_name
FROM
      customers
WHERE
      last_name LIKE 'Ch%';
```

You create a SQL query similar to below, where <> means "does not equal":

```
SELECT
      *
FROM
      Employee
WHERE
      jobCode <> 'INT'
      AND salary <= 30000;
```

Resources to learn more

- https://www.w3schools.com/sql/default.asp
- https://towardsdatascience.com/sql-cheat-sheet-776f8e3189fa

Data visualization

Planning a data visualization

Steps to plan a data visualization

Step 1: Explore the data for patterns

First, you ask your manager or the data owner for access to the current sales records and website analytics reports. This includes information about how customers behave on the company's existing website, basic information about who visited, who bought from the company, and how much they bought.

While reviewing the data you notice a pattern among those who visit the company's website most frequently: geography and larger amounts spent on purchases. With further analysis, this information might explain why sales are so strong right now in the northeast—and help your company find ways to make them even stronger through the new website.

Step 2: Plan your visuals

Next it is time to refine the data and present the results of your analysis. Right now, you have a lot of data spread across several different tables, which isn't an ideal way to share your results with management and the marketing team. You will want to create a data visualization that explains your findings quickly and effectively to your target audience. Since you know

your audience is sales oriented, you already know that the data visualization you use should:

- Show sales numbers over time
- Connect sales to location
- Show the relationship between sales and website use
- Show which customers fuel growth

Step 3: Create your visuals

Now that you have decided what kind of information and insights you want to display, it is time to start creating the actual visualizations. Keep in mind that creating the right visualization for a presentation or to share with stakeholders is a process. It involves trying different visualization formats and making adjustments until you get what you are looking for. In this case, a mix of different visuals will best communicate your findings and turn your analysis into the most compelling story for stakeholders. So, you can use the built-in chart capabilities in your spreadsheets to organize the data and create your visuals.

There are many different tools you can use for data visualization.

- You can use the visualizations tools in your spreadsheet to create simple visualizations such as line and bar charts.
- You can use more advanced tools such as Tableau that allow you to integrate data into dashboard-style visualizations.
- If you're working with the programming language R you can use the visualization tools in RStudio.

Week-5

Data analyst job opportunities

Fairness

- Fairness means ensuring that your analysis doesn't create or reinforce bias.
- In other words, as a data analyst, you want to help create systems that are fair and inclusive to everyone.

Some important points

➔ An issue is a topic or subject to investigate. (overall question)
➔ A question is designed to discover information. (question from the the overall question)
➔ A problem is an obstacle or complication that needs to be worked out.
➔ (problem in overall question)

Exploring your next job

Data analyst roles and job descriptions

Decoding the job description

To name a few others that sound similar but may not be the same role:

- **Business analyst** — analyzes data to help businesses improve processes, products, or services
- **Data analytics consultant** — analyzes the systems and models for using data
- **Data engineer** — prepares and integrates data from different sources for analytical use

- **Data scientist** — uses expert skills in technology and social science to find trends through data analysis
- **Data specialist** — organizes or converts data for use in databases or software systems
- **Operations analyst** — analyzes data to assess the performance of business operations and workflows

Decoding the job description

	Data Analysts	Data Scientists	Data Specialists
Problem solving	Use existing tools and methods to solve problems with existing types of data	Invent new tools and models, ask open-ended questions, and collect new types of data	Use in-depth knowledge of databases as a tool to solve problems and manage data
Analysis	Analyze collected data to help stakeholders make better decisions	Analyze and interpret complex data to make business predictions	Organize large volumes of data for use in data analytics or business operations
Other relevant skills	• Database queries • Data visualization • Dashboards • Reports • Spreadsheets	• Advanced statistics • Machine learning • Deep learning • Data optimization • Programming	• Data manipulation • Information security • Data models • Scalability of data • Disaster recovery

Other industry-specific specialist positions that you might come across in your data analyst job search include:

- *Marketing analyst — analyzes market conditions to assess the potential sales of products and services*
- *HR/payroll analyst — analyzes payroll data for inefficiencies and errors*
- *Financial analyst — analyzes financial status by collecting, monitoring, and reviewing data*
- *Risk analyst — analyzes financial documents, economic conditions, and client data to help companies determine the level of risk involved in making a particular business decision*
- *Healthcare analyst — analyzes medical data to improve the business aspect of hospitals and medical facilities*

Quick Review

Week-1

- Case study
- Steps in data analysis
  - ❖ **Ask**
  - ❖ **Prepare**
  - ❖ **Process**
  - ❖ **Analysis**
  - ❖ **Share**
  - ❖ **Act**
- Data and gut instant

Week-2

Analytical Skills

- **Curiosity**
- **Understanding context**
- **Having a technical mindset**
- **Data design**
- **Data strategy**

Analytical Thinking

- **Visualization**
- **Strategy**
- **Problem-orientation**
- **Correlation**
- **Big-picture and detail-oriented thinking**

Week-3

- **Data life cycle**
- **Data analytics tools**

Eg: Spreadsheet, Database, Visualization tools including Looker and tableau.

Week-4

- **Spreadsheet**
- **SQL**
- **Data visualization**
  - ❖ Tableau
  - ❖ Spreadsheet charts
  - ❖ R studio or Python

Week-5

- **Jobs for data analyst**
- **Different types of jobs**

Dhamodharan
28/09/2021