</talentlabs>

CHAPTER 5

Course Conclusion

</talentlabs>

AGENDA

- Course Summary
- Progression Advice
- Final Project

Course Summary

</talentlabs>

Data Ecosystem Structure

- Internal Data Sources
- External Data Sources

- Automatic Data Collection
- Manual Data Collection

- Joining datasets
- Cleaning data
- Handling missing values
- Renaming columns
- Deleting data

- Statistical methods
- Data Visualisation

- Accessible
- Secure

Important to note:

- Components interact
- Emerging technologies (Cloud Computing, Machine Learning and Big Data) are reshaping the data ecosystem

</talentlabs>

What are the key roles within the data ecosystem?

Data Engineer

- Organise Data
- Maintain Data Architectures

Data Analyst

- Generate Insights
- Organise data to draw conclusions

Data Scientist

- Combine Data Engineering and Data Analytics as well as Machine Learning

Business Analyst

- Leverage predictions and insights from data scientists and data analysts

</talentlabs>

Data Analytics Workflow

1. **Understand the problem and the desired outcome**
2. **Data collection**
   - Acquiring data that best serves the problem
3. **Data wrangling**
   - Optimising data for the task
4. **Data analysis and visualisation**
   - Look for insights
5. **Document the process**
   - Ensure process can be repeated
6. **Communicate insights to stakeholders**

➢ In this course we'll go through the entire workflow

Software

- Spreadsheets
  - Easy to learn
  - Well documented

Google Sheets

- Python
  - Open source with powerful libraries (NumPy, Pandas and Matplotlib)

- Tableau
  - Interactive data visualisation
  - Fast
  - Easy to use
  - Drag and drop features
  - Works well with huge datasets

Google Analytics is a free tool developed by Google that collects and compiles digital data into useful reports.

Google Analytics

Universal Analytics (UA)

- Older (beta 2012)
- Many websites still use UA
- Intuitive for beginners
- Ran in parallel with GA4 today
- Cookies based
- Sessions

Google Analytics 4 (GA4)

- Newer (beta 2020)
- AI capability
- Interaction based
- Events

This course

What is Tableau?

- Analytics platform and data visualisation tool
- Fast and easy to use
- No coding!
- Unique charts

tableau®

Excellent chart options in Tableau

- Bar Charts: for comparing categorical data
- Bubble Charts: for comparing quantity
- Maps: for presenting geocoded data
  - (also heatmaps: for showing relationships between two factors)
- Highlight Tables: for tabular presentation
- Bullet Graphs: to compare goals with performance
- Line Graphs: to investigate trend over time
- Scatter Plots: to understand relationships and trend

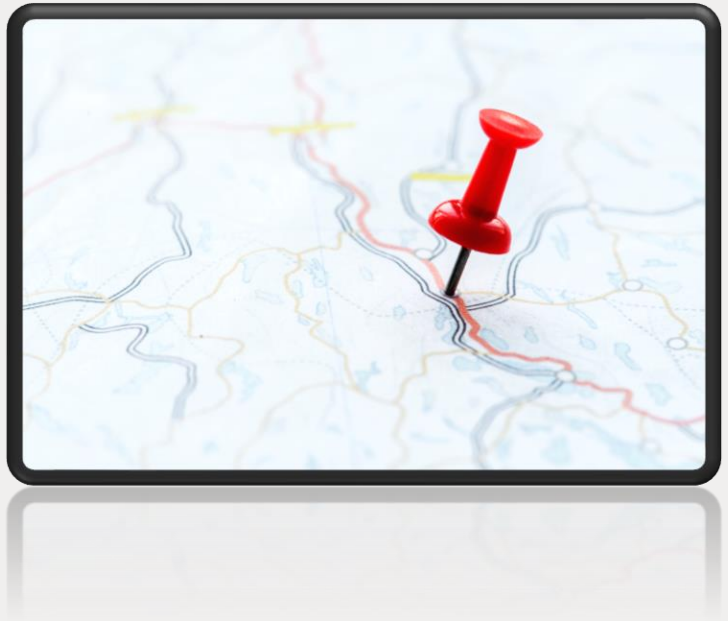
</talentlabs>

Best Visualisation Practices

- Story!
- Leverage visual elements:
  - Shape (mark shape, shape representation)
  - Colours (mark colour, colour palettes, colour representation, clarity)
  - Size (mark size, proximity)
  - Data to ink ratio (focus on essential visuals)
  - Labels (present, easy to read, clear)
  - Annotations (for additional detail)
  - Axes (multiple if needed, axis range)
  - Headlines (engaging?)

Dashboard Tips

- Think about the dimensions (mobile vs desktop)
- Add a title
- Add logo
- Utilize containers
- Update sheet specific tool tips
- Add Actions to make it interactive! (Dashboard > Actions…)
- Float filters (shift + left click to drag away from rigid format)
- Adjust filters (dropdown, multi-select, etc)
- Fill out the space (but think about the data-to-ink ratio)



Progression Advice

- Always learning!
- Check out Kaggle for free datasets to practice on
- Your own visualisation project!
- Medium articles
- Skills requirements

Final Project

</talentlabs>

Final Project

- Tests knowledge gained in the whole course
- Work through the entire data analytics workflow and apply the three tools we learned
  - Google Analytics (Data Collection)
  - Google Sheets (Data Wrangling)
  - Tableau (Data Visualisation)
- The end goal is to share a dashboard of your analysis

The Task

Your employer is interested in analysing the data from the Google Universal Analytics Demo Account. The report of interest is the Location report for the United States of America for the whole of 2021. The task is to analyse this data (via at least 4 charts) and to make a dashboard to showcase your analysis. It's up to you to decide which and how many columns of data to include in your analysis. Can you find something interesting?

The Task Part 1

01 Google Universal Analytics Demo Account

- Geo Location Report – America
- Set date range to the whole of 2021 (1st Jan 2021 – 31st December 2021)
- Select all rows
- Export

02 Import to Google Sheets

03 Prepare file for data wrangling (organise the file)

04 Data wrangling

Hints:

- Should have 50 states (clean duplicates + manually check for errors)
- Remove the Totals row
- Check every column for correct format (i.e. percent, time, number, etc)
- Apply Left and Right functions on revenue column

Google Analytics

Google Sheets

</talentlabs>

The Task Part 2

05 Data visualisation in Tableau

- Use charts of your choice (must include 1 map chart and have at least 4 charts in total)
- Make it fun! (Experiment!)
- Use a colour scheme
- Add a title to the Dashboard - Google Analytics United States 2021 Visualisation by Insert Name (e.g. John Smith)
- Can look for inspiration on Tableau Public Discover
- Don’t worry about interactivity (Actions etc) as file will be assessed as an image

06 Publish Dashboard

- Make it private
- Download image and submit
- Share the link in submission