Assignment Details

Course title

**NCFE Level 3 Certificate in Data (603/7882/7)**

Unit

**01 Understand how to source data (L/618/8650)**

Task

**End of unit assessment**

Tutor

**Komal Karir**

Learner

***Dennis Ideler***

Date submitted

***Mon 24 March 2025***

Candidate Statement

All work submitted in your portfolio must be your own.

If you copy from somebody else or allow another candidate to copy from you then you may be disqualified.

**I have read and understood the Notice for Candidates**

**I confirm:**

**• This submission is my own work**

**b) Any help from others, excluding the tutor(s), has been explained below.**

**Signature:** Dennis Ideler

**Date:** 2025-03-24

Notes:

• Please answer using full sentences. Make sure to check spellings.

• Please ensure you rename this file to include your name at the beginning.

Question 1.0: Understand where common sources of data can be found

**Outline how and where can you obtain the information required to prepare a report for the potential client. Include URLs to publicly available data services.**

**(You do not need to prepare the report yourself).**

*BACKGROUND:*

*Your business offers investment advice to short-term investors, specialising in tech companies listed on the London Stock Exchange (LSE).*

*The company stores the following data*

*- Customer personal details*

*- Customer financial details*

*- Customer tax details*

*- Customer investment portfolio*

*- Customer investment advice (i.e. notes from confidential meetings with investment staff)*

*- ... transactions, invoices, staff, companies, stocks, and much more...*

*Investment staff can access client limited client data, read-only, for giving investment advice.*

*Only authorised staff have write operations (create, update, delete) to client data.*

*Clients can log in to your website and access their portfolio data. They cannot yet submit deal instructions (i.e. buy, sell).*

*Investment recommendations are done via a generated report for the client. The report contains several companies with*

*- key staff profiles (e.g. executives)*

*- historic data (e.g. founded date, filings, dividends)*

*- past performance (e.g. financial statements, profit/loss)*

*- company ethos (e.g. green targets, no animal testing, human/employee rights)*

*- share price movements (past 6 months, though a real product would be much longer)*

*TASK:*

*A client wants to invest £10k. You have selected 5 tech companies as suitable investment opportunities.*

*Outline how and where you can obtain the information required to prepare the report. Include URLs.*

*ANSWER:*

*Some financial data services are only catered to the casual individual consumer and data is only available within their app (e.g. Apple Stocks).*

*This data would need to be scraped and not only is that inefficient, it is in violation of the End User License Agreement, so we will be ignoring such services.*

*Instead we'll focus on services that provide some form of API access to financial data so we can programmatically fetch the data.*

*This will most commonly be via HTTP REST APIs but does not need to be limited to that (e.g. FTP APIs would also be suitable).*

*Your company's web servers will run bespoke software that loads data stored in the database, fetches data from online services, analyses the data,*

*generates a report, saves the report, and presents the report.*

*Let's describe some of those steps in more detail. The codebase will contain connectors and adaptors for accessing data.*

*For relational data that lives in the company's database, a DBMS connector will connect to a database server. SQL queries against*

*DB tables will return results of data. For example, get the client's personal details (e.g. name and address), financial details (e.g. how much they want to invest),*

*their investment portfolio (e.g. which stocks and how many shares do they own), their preferences and investment advice (e.g. what's important*

*to them, what's their risk appetite, do they want growth or cash flow, etc). It will also connect to various APIs over the network to fetch relevant data.*

*Given a list of companies that match the user's investment preferences and recommendations, it will collect relevant data about these companies.*

*Such as key staff, historic data, past performance, company ethos, share price movements. Note that fetching data from different sources can be done in parallel.*

*This data will be analysed and made to be presentable. Graphing and plotting libraries will be used at this point.*

*The generated report can be programmatically output in various formats (e.g. web page, email, PDF).*

*Some of these assets can be stored for later retrieval (e.g. put the generated PDF in an AWS S3 bucket).*

*The company may decide to have an optional step for reviewing and approving the generated report, before it's shown to the user.*

*Sample online data sources:*

*- Name: Companies House*

*- URL: https://find-and-update.company-information.service.gov.uk (web form), https://developer.company-information.service.gov.uk/get-started (REST API)*

*- Why: Access incorporation date, company status, nature of business (SIC code), registered office address, filings (accounts and confirmation statements), directors/officers (past and present)*

*- Name: Simply Wall St*

*- URL: https://api.simplywall.st/docs/*

*- Why: Access exchange data, company data, historical, pricing information, dividends, company insights and much more - and all for free while in beta!*

*- Name: yfinance (Yahoo! Finance)*

*- URL: https://yfinance-python.org/reference/index.html*

*- Why: Access market data and news about companies, historical and daily stock price data (do not need real-time data for this report)*

*- Name: London Stock Exchange Group*

*- URL: https://developers.lseg.com/en/api-catalog*

*- Why: Access trade information direct from the source (most of it will be overkill and overpriced for our needs, but adding to the list for completeness sake)*

*- Name: Investor relations company pages*

*- URL: Varies on the company recommend (e.g. https://investor.atmeta.com, https://ir.tesla.com, https://wise.com/owners, https://www.sage.com/investors, ... NB some of these companies are not listed on the LSE, included as representative examples of Investor Relations pages)*

*- Why: Access any investor data that the company self-publishes (e.g. company info and ethos/values/principles, press releases, quarterly earnings, annual meetings, annual reports, governance documents, dividends, ...)*

*Noteworthy mentions of APIs I found that have some overlap but might be more geared towards the US market: https://www.alphavantage.co, https://twelvedata.com, https://marketstack.com*

Question 2.0: Understand data formats

**What tools, charts and diagrams could you use to explain your reasons for recommending the five tech investments?**

Time-series line chart for stock price over a historical time range (e.g. 6 months). Can do one chart per company or show all companies on one chart for an easier comparison. Can also show a line for an index such as FTSE 100 as a comparison.

Bar chart for your personal performance, can show historical return rate grouped by month, and perhaps some estimated returns for future months. With a note that it's for illustrative purposes only.

Spider chart with edges for Value, Dividend, Health, Past, and Future to assess the overall status of a recommended company for investment.

Pie chart for revenue breakdown, can easily see income sources and how big of a slice they contribute. Similarly for expenses at a high level (e.g. R&D, marketing, production).

Question 3.0: Purpose and Functions of Data Architecture

**Considering your company’s database system, list 3 entities for which data must be held.**

**Prepare a brief table of the type of information (i.e. attribute) you would store for ONE of the entities you mentioned above. For each attribute define its data format (e.g. client’s first name, 20 characters, string)**

**How would you ensure clients’ information is secure from unauthorised access and maintenance?**

*The company would have many tables in its database for the various entities for which data must be held. For example*

*1. USERS could store personal details such as name and dob, and basic account information such login and security data.*

*Note that USERS could also be called CLIENTS or INVESTORS or CUSTOMERS.*

*2. PORTFOLIOS could store the investment holdings of users. Such as the shares they own, the number of shares, at what price they bought them, when they bought them.*

*Note that PORTFOLIOS could also be called USER\_PORTFOLIOS or INVESTMENT\_PORTFOLIOS or USER\_HOLDINGS.*

*3. INVESTMENT\_ADVICE could store the confidential meeting notes that clients have with investment staff. Such as the meeting minutes, date and time, location of the meeting, who was present, recommendations.*

*Note that INVESTMENT\_ADVICE could also be called CLIENT\_MEETINGS or CLIENT\_MEETING\_NOTES.*

*Let's define the USERS table in more details with a SQL-like pseudocode.*

*CREATE TABLE users (*

*id BIGINT PRIMARY KEY AUTO\_INCREMENT,*

*email VARCHAR(255) NOT NULL UNIQUE,*

*password\_hash VARCHAR(255) NOT NULL,*

*password\_salt VARCHAR(255) NOT NULL,*

*first\_name VARCHAR(100) NOT NULL,*

*last\_name VARCHAR(100) NOT NULL,*

*date\_of\_birth DATE NOT NULL,*

*phone\_number VARCHAR(20),*

*created\_at TIMESTAMP WITHOUT TIME ZONE NOT NULL DEFAULT NOW(),*

*updated\_at TIMESTAMP WITHOUT TIME ZONE NOT NULL DEFAULT NOW()*

*)*

*Instead of having a very wide USERS table, can break down user data over various tables such as USER\_ACCOUNTS, USER\_TAX\_DETAILS, USER\_ADDRESSES, USERS\_KYC, USER\_SETTINGS, and so on where the user\_id is a common foreign key.*

*Ensuring the clients' information is secure from unauthorised access and maintenance requires several approaches of security.*

*- The databases will have authorisation roles with various permissions (e.g. read-and-write, read-only, restricted tables) and authentication (username and password).*

*- Only the servers have read/write access to the primary database, and the application layer above manages access requests.*

*- Users can access their own data through the web application, which only fetches data for the currently logged in user. It may also allow them to write a limited set of data (e.g. add/update/delete an address).*

*- Sensitive data will be stored encrypted, such as passwords. For extra security, password hashes will also use salts.*

*- User authentication will be multi-factor to ensure only the actual user has access, e.g. email/password + security questions + one-time-password.*

*- Sessions will time-out after a set amount of time or inactivity, after which the user has to reauthenticate. This is to reduce the chance that someone can access a logged in account on an open computer for example.*

*- Special staff that need access to company and user data for their work will access a read-only replica database. All operations will be stored in an audit trail.*

*- Database backups are stored encrypted at rest.*