

Senior Data Engineer assignment

Let's do a quick end-to-end project which mimics the work that is expected from a Senior Data Engineer at Qure.ai. You will develop an Extract-Transform-Load (ETL) application to get key insights from the data. So please figure out all the compatible tools/integrations before working on the assignment. You have to use python to write this application. You can take up to 5 days to finish the assignment. Please send your submission in the form of Github or a pdf report with screenshots (please attach screenshots/link of dashboard and not the dashboard file). You will be asked to explain your solution during the next round.

Cleaning and storing the data [30 points]

You have one csv file containing various covid related data of US states starting from month of January 2020. Choose the best way/structure to store the data in any database (preferably a relational database like Postgres).

Visualisation and key insights [40 points]

Connect the database to any tool like Power BI or Tableau and display all the data in a meaningful way using dashboards (this can be bar charts, line graphs). There should also be relevant filtering and sorting options (for selecting state, date range, etc.)

One example can be the Case Fatality Rate (CFR) variation across states in US.

Sending reports [30 points]

Write an automated script that sends a periodic data summary to Slack (preferred) or to an Email address. You can get a free slack account and establish a connection with the database to post the summary. For this exercise, you can post the monthly trend analysis of the number of covid deaths from top 3 states. You can be creative and figure out some

different mechanism (other than slack/email) as well - the goal is to send updates to a group of users about important metrics periodically.

The slack message or email from bot should look like this:

Top 3 states with the highest number of covid deaths for the month of March

Month - March

State #1 (eg. NY) - no. of deaths, % of total US deaths

State #2 - no. of deaths, % of total US deaths

State #3 - no. of deaths, % of total US deaths

There should be 4 messages posted at a fixed interval for the month of March, April, May and June.