Time limit: 5 hours

Please read the entire document before starting. Results should be submitted in the form of an R Notebook (.Rmd file) with the compiled html file also attached. Email the results [nitish.goyal@welligence.com](mailto:nitish.goyal@welligence.com) with subject line:

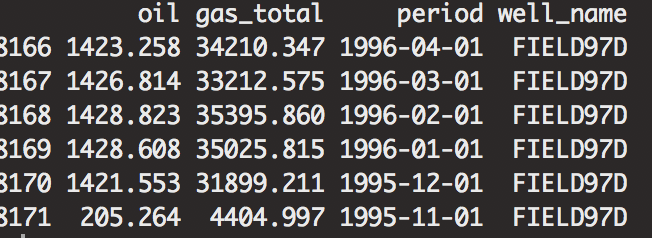
Rohit Tananki: Completed Technical Interview 04/13/2024.

Input: test.csv

* Columns: oil, gas, period, well\_name

Each row of the data frame represents a month of production for a specific well in the field.

For example the last row below indicates that well FIELD97D produced 205.264 barrels/d of oil in the month of November in 1995, and 4404.97 m3/d of gas.



**Data Cleaning and Exploration:** *tidyverse/dplyr*

1. produce a “series” data frame where the columns are well names and the rows are the date (period)
2. Create a data frame of well characteristics, where for each well in the data frame

The average gas to oil ratio is calculated, the number of months of production, the

Initial (first month date) of oil production, and the average month over month decline rate for

Gas.

**Forecasting**

* Filter out wells with < 24 months of production, filter out zero production.
* Using the filtered data, build and validate a time series forecasting model that predicts oil production. The time horizon for the prediction is 6 months out into the future. After trying a number of approaches, present and justify the best model. Note: In this part of the interview you are being asked to fit a model using all of the available well data with > 24 months of production, not to forecast a single well time series.

The solution to this part should include both the R code, and a paragraph detailing the methods tried, including any exploratory data analysis or modelling choices, and summarizing your findings.

**Visualization**

Write a function that takes as an input a well name, and outputs a plot of the well historical oil production values, along with the forecasts from your model (in a different color), where the x-axis is months since start of production.

**PostgreSQL**

Given the below database schema, write SQL which outputs a table with the following columns:

asset\_name (this is the name field in the assets table)

basin (this is the basin name from the basins table that corresponds to the asset)

pdp\_oil (from the asset\_reserves table)

average\_oil\_production (calculated as the average of the oil for each asset from the asset\_total\_productions table)

order the table by the value of the average\_oil\_production

A close up of text on a white background

Description automatically generated

**Other R Experience**

**Do you have any experience building RShiny apps? If so please detail it below.**

I have experience building interactive web applications with RShiny. During my experience at Edunovus , I developed a dashboard that enabled users to interactively visualize sales data over time. Users could filter the data by region, product category, and time period. I integrated ggplot2 for the visualizations and used RShiny’s reactivity system to update charts and tables based on the input.

**Do you have experience collaborating on github?**

Yes, I frequently use GitHub for version control and collaboration. I have worked on several projects where we used GitHub for issue tracking, code reviews, and managing pull requests. I am comfortable with common Git commands and have experience using GitHub features like branching and merging to manage project development collaboratively.

**What experience if any do you have deploying R in production?**

While I have not had the opportunity to deploy an R-based model in a production setting, I possess a strong foundation regarding the deployment of statistical models in production environments. I am knowledgeable about crucial aspects such as model validation, writing clean and maintainable code, ensuring reproducibility of results, and establishing comprehensive logging and monitoring systems. Alongside .I have experience deploying machine learning and NLP models built in Python to production using Google Cloud Platform (GCP) . I am enthusiastic about leveraging my Python deployment experience and my R knowledge in a practical setting.