Data Analyst II, III, Sr. – #12485

Presentation for Interview

Problem Statement:

Chelan PUD is seeking a Data Analyst to design, develop, enhance and utilize data analytics, business intelligence, and business processes to build and sustain a 'best in industry' data-driven decision-making environment, supporting a culture of continuous learning.

As part of the selection and interview process for this position, we would like you to demonstrate your abilities by manipulating and analyzing the attached data set, develop solutions for the listed questions, and present your findings to the interview team in the form of a PowerPoint. You will have up to 20 minutes for the presentation.

Data Set Background:

This Data Analyst role will interact with many different departments across the District, with unique data, reporting and analytics needs. Attached is a data set from one of our constituents, Hydro Generation, representing the small generation units for Rocky Reach Dam for the year 2018.

The data are hourly averages, and the associated timestamp is for the start of the hour (i.e. the row labeled 3:00 am is for the hour from 3:00am – 4:00am). Each row includes two timestamps, one in UTC Time and one in Pacific Time.

Do not assume the data are perfectly clean. If you do any extra data cleaning, please explain what you have done and why.

For Rocky Reach Small Units C2, C4, C5, C6, and C7, these are the measurements:

- avg_winding_temp(C) – average stator winding temperature in degrees Celsius
- total_current(A) – total generator electrical current in amperes
- avg_cooling_water_flow(gal/min) – average cooling water flow in gallons per minute
- avg_cooling_water_temp(C) – average cooling water temperature in degrees Celsius
- avg_cooling_air_out_temp(C) – average cooling air temperature as it leaves the cooler and returns to the stator in degrees Celsius

Questions to be Addressed in the Presentation:

1. How many NA errors are there in the dataset?

   Special notes:

   - *For simplicity, and to answer the rest of the questions, fill all NA values with 0.*

   - *Amperes/amps, are a measure of electrical current. For consistency, and due the nature of sensor data, call a unit 'offline' if it has an hourly average total current less than 10A*

2. How many total hours was Unit C6 offline in 2018?

3. How many hours was C6 offline in each month?

4. What percentage of the time was C6 offline in each month?

5. Which unit ran for the most amount of time in 2018?

6. Which hour of the day has the most offline time?

7. Create a graph showing the average cooling water temperature by month. Are there any other data points that have seasonality fluctuations?

8. Are there any apparent differences between the cooling systems of the 5 units? Show the difference graphically. (I.e. Do any of the units run hotter?)

9. For the month of May, design a dashboard for the cooling system that shows some graphs and relevant statistics (max, min, average, etc.) to display how the cooling system performed during the month.

10. What other type(s) of analysis or business intelligence reporting could be done to help understand the cooling systems and provide insights into their operational performance. What software tools and methods would you use?

Reference Information: Generator Stator, Rotor, and Cooling System

Although not required reading to answer the previous questions, you may find the following information, and operations and maintenance information useful in understanding the fundamentals of generating unit at Rocky Reach Dam:
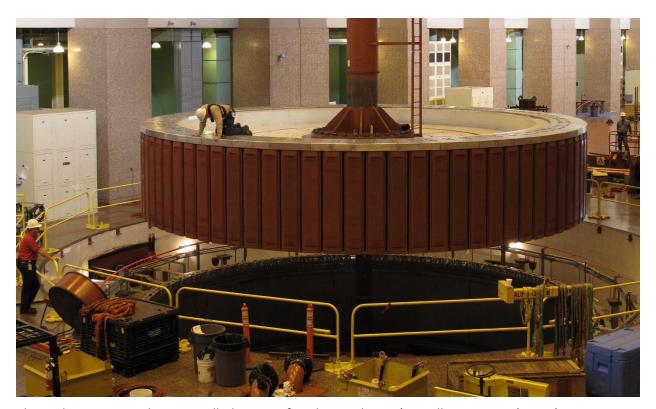

Photo above: A rotor being installed in one of Rocky Reach Dam's small generators (C1-C7).

PDF of Generator Rotor and Stator:

RotorAStator_R...

PDF of Generator Cooling Water System:

GenCoolingWat...