

Exploratory Data Analysis Assignment

Centrifugal Pump Sensor Data Analysis

Sonu Yadav, Sagnik Halder, Sachin Singh, Ravi

1 Introduction

You have been provided with sensor data collected from a centrifugal pump over an extended period of operation. Your task is to perform exploratory data analysis (EDA) to extract meaningful insights from this data.

2 Datasets

You will work with two datasets:

Dataset 1: Clean Pump Data

- **Link:** <https://drive.google.com/file/d/1b7GrsmovxbJa-430mE2c203PgItDCDp6/view?usp=sharing>
- **Use for:** Questions 1 and 3

Dataset 2: Renamed Sensor Data

- **Link:** https://drive.google.com/file/d/1S96u5gXLHQDMCEYGCwPNZhSU7jgi_vb0/view?usp=sharing
- **Use for:** Question 2

Important Note

The datasets contain measurements from various sensors monitoring different aspects of pump operation including speeds, temperatures, pressures, vibrations, and flow rates.

3 Question 1: Operating Mode Detection

Industrial pumps often operate at discrete speed settings rather than continuous variable speeds. Using the **Clean Pump Data**, analyze the **Pump Shaft Speed** column.

Question

How many distinct speed settings does the pump motor have?

Use appropriate visualization and analysis to identify the number of operating modes and the approximate RPM value for each mode.

4 Question 2: Redundant Sensor Identification

In industrial settings, multiple sensors may measure the same physical quantity for redundancy. Using the **Renamed Sensor Data** (columns labeled A-Z), identify sensor relationships through visual analysis.

Question - Part A

Which sensors are measuring the same kind of data?

Identify groups of sensors that show nearly identical patterns.

Question - Part B

Which pair of sensors shows an inverse (negative) relationship?

Identify at least one pair of sensors where when one increases, the other decreases proportionally.

Important

Keep your findings from this question in mind — they will be crucial for Question 3!

5 Question 3: Feature Selection for Predictive Modeling

Your team plans to build a regression model to predict Pump Radial Bearing Vibration. Before model training begins, you need to perform EDA to understand the feature landscape.

Part A: Feature Elimination via EDA

Using the **Clean Pump Data**, analyze all available features in relation to Pump Radial Bearing Vibration.

Question

Which features should be eliminated and why?

Through EDA, identify and list the names of features that should be removed. Provide justification for each elimination.

Critical Insight

Connect the dots: Use your Question 2 findings to inform your feature selection strategy!

Part B: Constrained Feature Selection

Suppose the model training system has a strict constraint and can only accept exactly **4 features** as input.

Question

Which 4 features should one use?

Based on your EDA, recommend exactly 4 features. List the 4 feature names with supporting evidence from your analysis.

Good luck with your analysis!