Analysis of Data Using Python and Power BI

TATA STEEL GLOBAL WIRES LTD.

TARAPUR MIDC

Prepared By:

Under the guidance of

Mr. PRASHANT AROTE
ITS HEAD

Mr. RUPESH CHANDRA
MANAGER ITS

Global Wires

Tata Steel Limited

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Place: Tarapur, MIDC
Date: 14-07-222

Abstract

Big data not only means a huge amount of data but also a data which is high in variety. Recent times have seen a massive jump in the amount of data available. It is difficult to gain valuable insights and knowledge from the datasets by traditional techniques. The rising trend of data science in the past years has also meant a sharp rise in data analysis techniques. Modern tools and techniques save a lot of time in hefty processes and allow to finish the work of days in minutes. This paper aims to analyze some of the different analytics methods and tools which can be applied to big data, as well as the opportunities provided by the application of big data analytics in various decision domains. Data Analysis is process of inspecting, cleansing, transforming, and modelling data to discover useful information, informing conclusions, and supporting decision-making. Data Analysis can be broadly classified into Quantitative Data Analysis and Qualitative data analysis. EDA focuses on discovering new features in the data while CDA focuses on confirming or falsifying existing hypotheses. Predictive analytics focuses on the application of statistical models for predictive forecasting or classification, while text analytics applies statistical, linguistic, and structural techniques to extract and classify information from textual sources, a species of unstructured data.

Keywords: Data analysis, processing, cleaning, extraction, standard deviation.

Table of Contents

Chapter 1

Introduction

Imagine a world without data storage; a place where every detail about a person or organization, every transaction performed, or every aspect which can be documented is lost directly after use. Organizations would thus lose the ability to extract valuable information and knowledge, perform detailed analyses, as well as provide new opportunities and advantages. Anything ranging from customer names and addresses, top-products available, top-purchases made, to employees hired, etc. has become essential for day-to-day continuity. Data is the building block upon which any organization thrives. Now think of the extent of details and the surge of data and information provided nowadays through the advancements in technologies and the internet.

1.1 Motivation

Nowadays, rapid change is observed in data in variety of view. Big data analytics is required along with different storage and analysis methods for the data to be properly analyzed and gain the required information. Data analytics has become increasingly important in the enterprise as a means for analyzing and shaping business processes and improving decision-making and business results. The aim of data analysis is to apply statistical knowledge and technologies on the big data to find trends and solve problems

1.2 Problem Statement

Analyzing Coils Data of Pickling Line. The objective of this project is to determine the best and the not ok coil among all the coils present in data stored for 1 month. By writing Python Modules to inspect, clean & transform data, it is possible to make estimation of a possible range which is best to run a coil. It has been taken into account the fact that a coil may not always run in the given standard values for the respective parameters. This analysis helps to determine highly increasing or decreasing values present in the data.

1.3 Objectives

The main objective of data analysis is to arrange the constantly increasing data ranging from its highest value to its lowest value in organize manner to derive valuable information. Organizing data aids to many statistical approaches and analyzation of the information through various viewpoints. To extract information in order to utilize the information gained to make better decisions. To implement various data analysis techniques like Exploratory Data Analysis, Confirmatory Data Analysis etc.

Chapter 2

Review of Literature

2.1 Study of data analysis model based on big data technology

The traditional data analysis is based on the cause and effect relationship, formed a sample microscopic analysis, qualitative and quantitative analysis, the thinking mode of trend extrapolation analysis. Big data has a fundamental impact on the traditional data analysis. Big data analysis based on correlation, formed global macro analysis, data and technical analysis, correlation analysis and new thinking mode of correlation analysis. Namely, from causal analysis to correlation analysis and knowledge discovery, from model fitting to data mining, from logical reasoning to association rules. Data analysis in the era of big data have taken great changes, Namely, Big data analysis, from the analysis of objects, the mode of data processing, analytical methods and tools, analytical thinking.

2.2 Some key problems of data management in army data engineering based on big data

This paper analysed the challenges of data management in army data engineering, such as big data volume, data heterogeneous, high rate of data generation and update, high time requirement of data processing, and widely separated data sources. We discussed the disadvantages of traditional data management technologies to deal with these problems. We also highlighted the key problems of data management in army data engineering including data integration, data analysis, representation of data analysis results, and evaluation of data quality.

2.3 Research on Big Data Analysis Data Acquisition and Data Analysis

Using big data analysis algorithm, this paper discusses the source of data acquisition, points out the technical characteristics of data source, and explains the data acquisition methods and requirements. Clear the purpose of big data analysis, build data analysis system, describe the process of data analysis. There are four steps in big data analysis: data acquisition, data storage, data analysis and data mining. Data acquisition can be divided into two parts: acquisition and preprocessing, which is just narrow data acquisition. The data analysis of the Internet of things is a process of organizing and purposefully collecting data, processing data and analyzing data to form information. The analysis process is the support and execution process of the algorithm. In the data processing cycle, data analysis should be used properly in every link from data acquisition in the field perception layer to data transmission in the communication layer to data processing in the application layer, so as to improve the effectiveness and rationality of data processing.

Chapter 3

Data Analytics

3.1 Characteristics of Big Data

Organizations today are facing problems of data organization. They have a wealth of information but are unable to extract valuable information out of it as the information they have is in its most raw form. The information or the data they have is either unprocessed, partial processed or unstructured format.

The term “Big Data” is applied to datasets where data shows much variations and grows so rapidly that is not possible to study and organize them using traditional tools and techniques. Big data sizes are constantly increasing, currently ranging from a few dozen terabytes (TB) to many petabytes (PB) of data in a single data set. Consequently, some of the difficulties related to big data include capture, storage, search, sharing, analytics, and visualizing.

Volume, Variety and Velocity are the three characteristics that define Big Data.

3.2 Data Analysis Techniques

- **Regression analysis**

Regression analysis is a type of statistical analysis method that determines the relationships between independent and dependent. experiments that involve manipulating the values of independent variables, a quantitative data analyst can assess the impact of the changes on the dependent variable.

- **Simple linear regression analysis**

A simple linear regression analysis formula includes a dependent variable and an independent variable. The mathematical representation of the dependent variable is typically Y, while X represents the independent variable.

- **Hypothesis analysis**

Hypothesis analysis is a data analysis technique that uses sample data to test a hypothesis. Hypothesis analysis is a statistical test method to validate an assumption and determine if it’s plausible or factual. In this approach, an analyst develops two hypotheses — only one of them can be true. Components of hypothesis analysis are the null hypothesis and the alternative hypothesis.

3.3 Data Analysis Tools

1. R and Python
2. Microsoft Excel
3. Tableau
4. RapidMiner
5. KNIME
6. Power BI
7. Apache Spark
8. QlikView
9. Talend
10. Splunk

3.4 Process in Data Analysis

3.4.1 ETL (Extract, Transform, Load)

ETL is a process that extracts, transforms, and loads data from multiple sources to a data warehouse or other unified data repository.

- **Extract**

During data extraction, raw data is copied or exported from source locations to a staging area. Data management teams can extract data from a variety of data sources, which can be structured or unstructured. The extraction sources include SQL servers, ERP systems etc.

- **Transform**

In the staging area, the raw data undergoes data processing. Here, the data is transformed and consolidated for its intended analytical use case. This phase can involve filtering, cleansing, de-duplicating, validating, and authenticating the data.

- **Load**

In this last step, the transformed data is moved from the staging area into a target data warehouse. Typically, this involves an initial loading of all data, followed by periodic loading of incremental data changes and, less often, full refreshes to erase and replace data in the warehouse. For most organizations that use ETL, the process is automated, well-defined, continuous and batch-driven. Typically, ETL takes place during off-hours when traffic on the source systems and the data warehouse is at its lowest.

3.4.2 EDA Process (Exploratory Data Analysis)

Exploratory Data Analysis is a method of evaluating or comprehending data in order to derive insights or key characteristics. EDA can be divided into two categories: graphical analysis and non-graphical analysis.

EDA is a critical component of any data science or machine learning process. You must explore the data, understand the relationships between variables, and the underlying structure of the data in order to build a reliable and valuable output based on it.

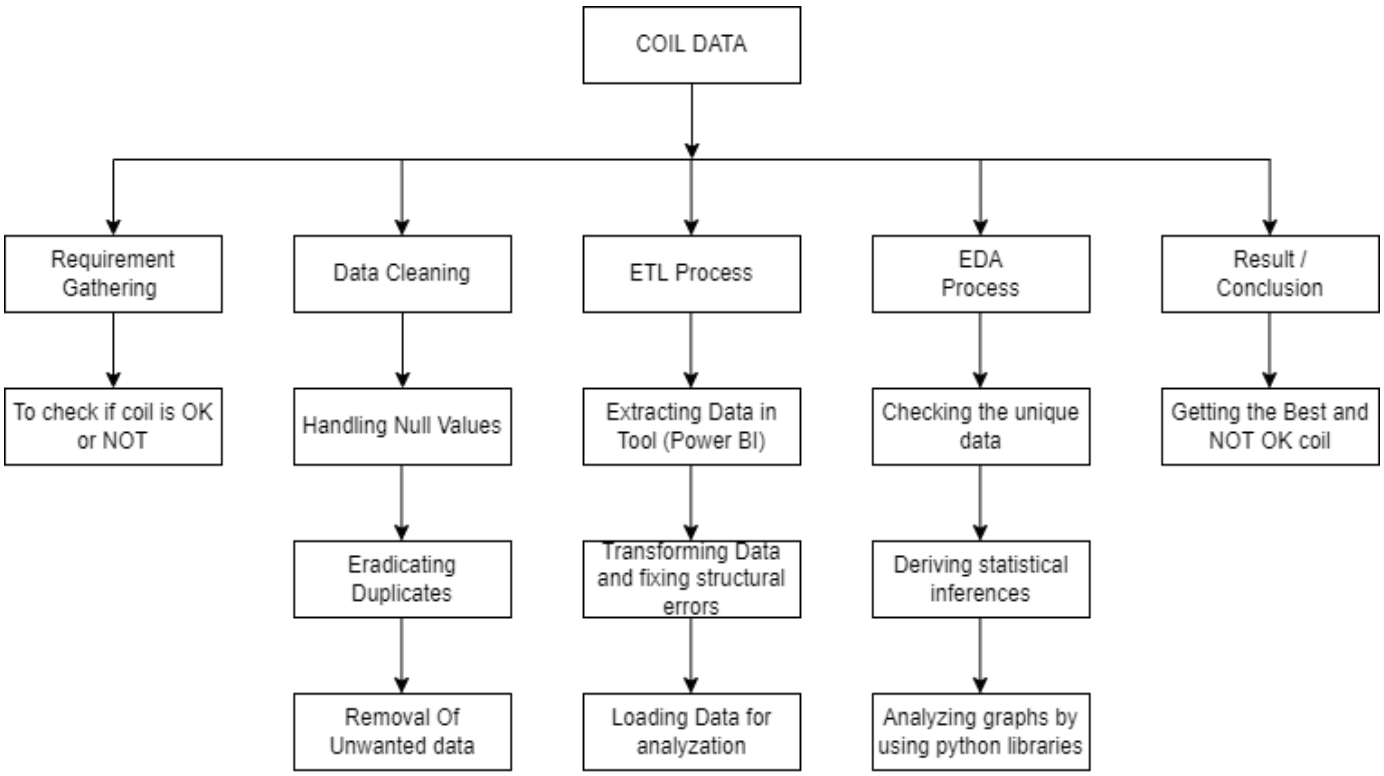
EDA Process includes:

1. Importing the Python Libraries
2. Loading the Dataset in Python
3. Structured Based Data Exploration
4. Handling Missing Values
5. Data Visualization

Chapter 4

Requirements Gathering and Planning

4.1 Work Breakdown Structure



4.2 Feasibility Study

4.2.1 Technical Feasibility

Technical Feasibility Assessment examines whether the proposed system can actually, be designed to solve the desired problems and requirements using available technologies. The system is said to be feasible technically if it can be operable and manageable under current technological context. Since the system aims to extract the valuable insights, via the use the data set which is available and the building a classifier is also feasible, the project can be considered technically feasible.

4.2.1.1 Hardware Requirements

- RAM: - 64 MB or above storage space 4 GB and more.
- Processor Speed: - 500 MHz and above
- Processor: -Above x86

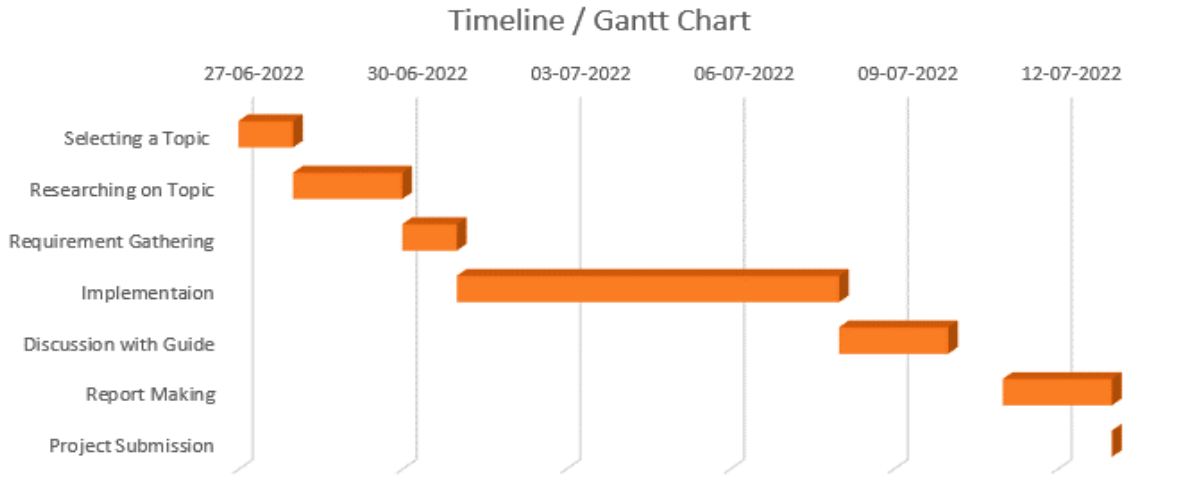
4.2.1.2 Software Requirements

- Web browser
- Windows 10
- Jupyter Notebook
- Microsoft Excel
- Microsoft Power BI
- Python version 3.9.5 and above

4.2.2 Economic Feasibility

Economic Feasibility checks whether the cost required for complete project development is feasible using the available resources in hand. It should be noted that the cost of resources should be kept minimum while operational and maintenance cost for the project should be within the capacity of organization. Since the system is hosted free of cost for limited use, the system can be considered economically feasible for the development.

4.3 Timeline and Gantt Chart



	Project Submission	Report Making	Discussion with Guide	Implementaion	Requirement Gathering	Researching on Topic	Selecting a Topic
Start	13-07-2022	11-07-2022	08-07-2022	01-07-2022	30-06-2022	28-06-2022	27-06-2022
Days to Complete	2	2	2	7	1	2	1

Chapter 5

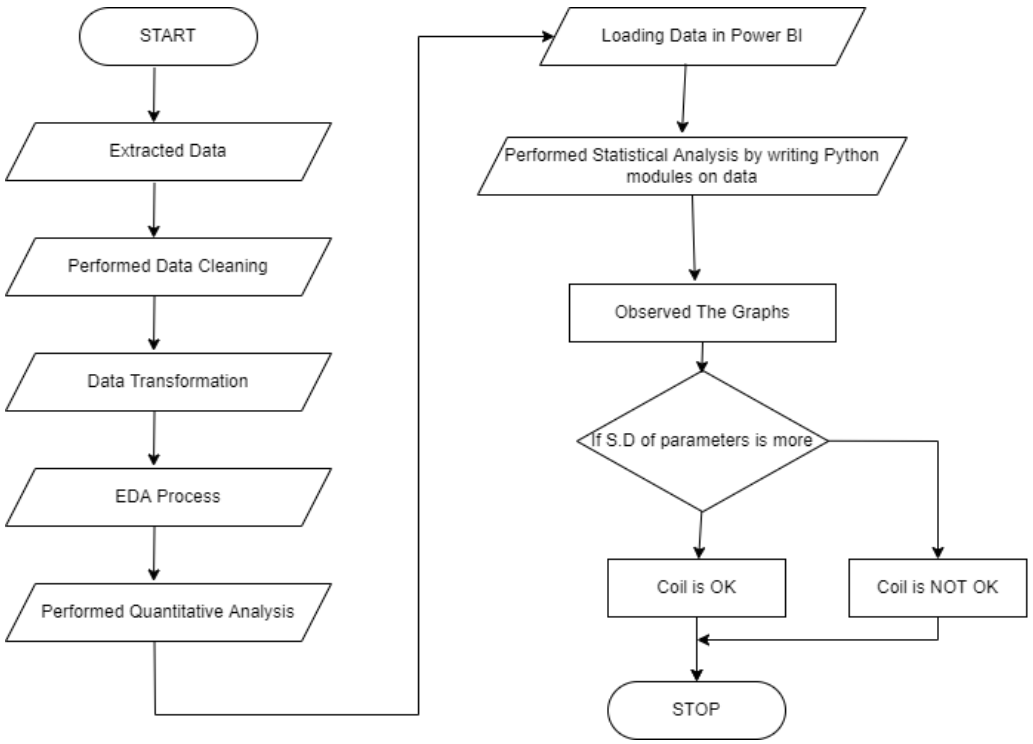
Report on Present Investigation

5.1 Proposed System

Iron raw materials when exposed to atmospheric gases and other chemicals in industry needs to undergo pickling processes. Iron coils when go through pickling observations are taken at various stages. From the data extracted for analysis, it is known that observations are made at every 10 seconds. Total time taken by the coil to undergo the process can be derived depending upon the no. of datapoints observed.

Considering various parameters like line speed of coil, coil segments, entry and exit in potentiometer, RC1; RC2; RC3 Temperature statistical analysis is to be done to find the best and the not ok coil. Along with this, analysis this data also helps find the range of values should lie for a specific coil to run in standard manner.

5.1.1 Flowchart of proposed system



5.2 Implementation

5.2.1 Dataset

The below screenshot contains the one-month data of Pickling Line and it contains the data of 1294 coils. The various parameters based on which the analysis has to be done are I_SEGMENTNR, I_COIL_NO, I_LINE_SPEED, I_DT_TIME, I_RC1_TEMPERATURE, I_RC1_TEMPERATURE, I_RC1_TEMPERATURE, I_LINEAR_POTENTIOMETER_ENTRY, I_LINEAR_POTENTIOMETER_EXIT.

	A	B	C	D	E	F	G	H	I	J
283758	26	19.88086128	99999	68.32559967	68.7508316	74.18092346	269.791687	257.6388855	13-06-2022 01:08	13-06-2022
283759	27	19.20550346	99999	68.33407593	68.73312378	74.02981567	269.7935181	257.6388855	13-06-2022 01:08	13-06-2022
283760	28	19.75808144	99999	68.37893677	68.69314575	73.87900543	269.791687	257.6388855	13-06-2022 01:08	13-06-2022
283761	29	19.87918091	99999	68.38990784	68.66717529	73.72712708	269.791687	257.6388855	13-06-2022 01:08	13-06-2022
283762	30	19.94428062	99999	68.40409851	68.64224243	73.57652283	269.791687	257.6388855	13-06-2022 01:09	13-06-2022
283763	31	16.34824371	99999	68.42560577	68.62381744	73.43578339	269.791687	257.6388855	13-06-2022 01:09	13-06-2022
283764	32	1.030360103	99999	68.43506622	68.58986664	73.29772949	269.791687	257.6388855	13-06-2022 01:09	13-06-2022
283765	33	1.095780134	99999	68.44879913	68.5517807	73.15974426	269.4060059	257.6388855	13-06-2022 01:09	13-06-2022
283766	34	2.196460009	99999	68.46092224	68.53404999	73.02754211	200.4084167	257.6388855	13-06-2022 01:09	13-06-2022
283767	35	5.872442245	99999	68.49536896	68.51245117	72.90090179	171.703476	219.8610229	13-06-2022 01:09	13-06-2022
283768	36	1.49435997	99999	68.49536896	68.49370575	72.77377319	166.7947998	179.8847198	13-06-2022 01:10	13-06-2022
283769	37	17.09728432	99999	68.501297	68.44995117	72.65743256	174.3920135	176.5625	13-06-2022 01:10	13-06-2022
283770	38	21.41483307	99999	68.51703644	68.42796326	72.54184723	176.7104187	176.5625	13-06-2022 01:10	13-06-2022
283771	39	36.78158569	99999	68.51851654	68.41053009	72.43671417	179.2022552	176.5625	13-06-2022 01:10	13-06-2022
283772	40	43.10900116	99999	68.51851654	68.38974762	72.32595062	179.253479	176.5612793	13-06-2022 01:10	13-06-2022
283773	41	40.52978897	99999	68.51851654	68.35731506	72.22041321	179.5008698	176.5625	13-06-2022 01:10	13-06-2022
283774	42	40.57991028	99999	68.52344513	68.32363129	72.11698914	179.7671051	176.5625	13-06-2022 01:11	13-06-2022
283775	43	40.57991028	99999	68.54166412	68.30305481	72.0152359	179.8710785	176.5625	13-06-2022 01:11	13-06-2022
283776	44	40.57991028	99999	68.55060577	68.28289795	71.91638947	179.9565887	176.5625	13-06-2022 01:11	13-06-2022
283777	45	40.57991028	99999	68.5719986	68.26235962	71.82160187	180.0347137	176.5625	13-06-2022 01:11	13-06-2022
283778	46	40.57991028	99999	68.61574554	68.21879578	71.73316956	180.0993042	176.5625	13-06-2022 01:11	13-06-2022
283779	47	40.57991028	99999	68.63425446	68.21065521	71.66641998	180.1113281	176.5625	13-06-2022 01:11	13-06-2022

.

.

	A	B	C	D	E	F	G	H	I	J
1	I_SEGMENTNR	I_LINE_SPEED	I_COIL_NO	I_RC1_TEMPERATURE	I_RC2_TEMPERATURE	I_RC3_TEMPERATURE	I_LINEAR_POTENTIOMETER_ENTRY	I_LINEAR_POTENTIOMETER_EXIT	I_DT_TIME	Date
2	0	47.43988037	0	70.18518066	68.80268097	75.61974335	148.656601	135.762146	23-05-2022 00:12	23-05-2022
3	1	47.43988037	0	70.18518066	68.79324341	75.49634552	148.7847137	135.7638855	23-05-2022 00:12	23-05-2022
4	2	47.43988037	0	70.17375183	68.77370453	75.35423279	148.7847137	135.7638855	23-05-2022 00:12	23-05-2022
5	3	47.43988037	0	70.16204071	68.7698822	75.21141052	148.7847137	135.7638855	23-05-2022 00:13	23-05-2022
6	4	47.43988037	0	70.16204071	68.7589035	75.12375641	148.7847137	135.7638855	23-05-2022 00:13	23-05-2022
7	0	23.42116165	0	69.7547226	68.02555847	73.93595123	148.4372406	135.344101	23-05-2022 00:23	23-05-2022
8	1	0.196920097	0	69.71169281	68.06750488	74.19342804	147.5585938	134.7222137	23-05-2022 00:24	23-05-2022
9	2	0	0	69.68851471	68.12743378	74.42717743	147.9061584	134.6890564	23-05-2022 00:24	23-05-2022
10	3	0	0	69.6733551	68.19657135	74.6455307	147.4508667	134.6108398	23-05-2022 00:24	23-05-2022
11	4	0	0	69.65322113	68.29161835	74.84629822	147.9922638	134.5486145	23-05-2022 00:24	23-05-2022
12	5	0	0	69.65277863	68.39294434	75.02023315	148.1770782	134.5486145	23-05-2022 00:24	23-05-2022
13	6	0	0	69.65277863	68.50125122	75.17328644	148.1770782	134.5486145	23-05-2022 00:24	23-05-2022
14	7	0	0	69.65277863	68.6267395	75.30289459	148.1770782	134.5486145	23-05-2022 00:25	23-05-2022
15	8	0	0	69.65277863	68.75298309	75.40259552	147.540451	134.3848114	23-05-2022 00:25	23-05-2022
16	9	0	0	69.65731812	68.89305878	75.47398376	147.222229	134.1984406	23-05-2022 00:25	23-05-2022
17	10	0	0	69.67592621	69.03421021	75.48634338	147.1660614	134.0289917	23-05-2022 00:25	23-05-2022
18	11	0	0	69.69300842	69.17706299	75.47622681	205.7789917	171.3781281	23-05-2022 00:25	23-05-2022
19	12	0	0	69.71192169	69.32402802	75.41750336	268.6765747	257.4589233	23-05-2022 00:25	23-05-2022
20	13	0	0	69.75860596	69.46951294	75.3309021	268.663208	257.6388855	23-05-2022 00:26	23-05-2022
21	14	0	0	69.79685211	69.61064911	75.21796417	268.663208	257.6388855	23-05-2022 00:26	23-05-2022
22	15	0	0	69.82900238	69.74629211	75.08087921	268.663208	257.6388855	23-05-2022 00:26	23-05-2022

5.2.2 Screenshot of output with Description

- Python libraries like NumPy, Seaborn, Pandas are imported and data is imported using python modules.

```
[2]: import numpy as np
     import seaborn as sns
     import pandas as pd
```

```
[4]: data = pd.read_csv("COMBINE.csv")
```

```
16]: data.head()
```

16]:

	I_SEGMENTNR	I_LINE_SPEED	I_COIL_NO	I_RC1_TEMPERATURE	I_RC2_TEMPERATURE	I_RC3_TEMPERATURE	I_LINEAR_POTENTIOMETER_ENTRY	I_LINEA
0	0	47.43988	0	70.185181	68.802681	75.619743	148.656601	
1	1	47.43988	0	70.185181	68.793243	75.496346	148.784714	
2	2	47.43988	0	70.173752	68.773705	75.354233	148.784714	
3	3	47.43988	0	70.162041	68.769882	75.211411	148.784714	
4	4	47.43988	0	70.162041	68.758904	75.123756	148.784714	

```
17]: data.tail()
```

17]:

	I_SEGMENTNR	I_LINE_SPEED	I_COIL_NO	I_RC1_TEMPERATURE	I_RC2_TEMPERATURE	I_RC3_TEMPERATURE	I_LINEAR_POTENTIOMETER_ENTRY	I_
283773	43	40.57991	99999	68.541664	68.303055	72.015236	179.871078	
283774	44	40.57991	99999	68.550606	68.282898	71.916389	179.956589	
283775	45	40.57991	99999	68.571999	68.262360	71.821602	180.034714	
283776	46	40.57991	99999	68.615746	68.218796	71.733170	180.099304	
283777	47	40.57991	99999	68.634254	68.210655	71.666420	180.111328	

- Unique values in the data are derived.

```
: data.nunique()

: I_SEGMENTNR                1140
  I_LINE_SPEED              29201
  I_COIL_NO                  1290
  I_RC1_TEMPERATURE        138833
  I_RC2_TEMPERATURE        127913
  I_RC3_TEMPERATURE        172782
  I_LINEAR_POTENTIOMETER_ENTRY  124809
  I_LINEAR_POTENTIOMETER_EXIT    72752
  I_DT_TIME                  46904
  Date                          33
  dtype: int64
```

- After Data cleaning process is done, it is checked if any null values are found in the data. Following screenshot depicts that no null values were found in the data.

```
[19]: data.isnull()
```

[19]:

	I_SEGMENTNR	I_LINE_SPEED	I_COIL_NO	I_RC1_TEMPERATURE	I_RC2_TEMPERATURE	I_RC3_TEMPERATURE	I_LINEAR_POTENTIOMETER_ENTRY	I_
0	False	False	False	False	False	False	False	
1	False	False	False	False	False	False	False	
2	False	False	False	False	False	False	False	
3	False	False	False	False	False	False	False	
4	False	False	False	False	False	False	False	
...	...	...	...	...	...	...	...	
283773	False	False	False	False	False	False	False	
283774	False	False	False	False	False	False	False	
283775	False	False	False	False	False	False	False	
283776	False	False	False	False	False	False	False	
283777	False	False	False	False	False	False	False	

283778 rows × 10 columns

- Statistical Analysis of data is done using Power BI. Standard Deviation of all the parameter for every unique coil is obtained.
- Standard Deviation of the data gives how much does the values differ from the its previous and the preceding value.
- Coil with the least standard deviation proves to be best coil while coil with highest standard deviation is found to be the Not Ok coil.

I_COIL_NO	Standard deviation of I_LINE_SPEED	Standard deviation of I_LINEAR_POTENTIOMETER_ENTRY	Standard deviation of I_LINEAR_POTENTIOMETER_EXIT	Standard deviation of I_RC1_TEMPERATURE	Standard deviation of I_RC2_TEMPERATURE	Standard deviation of I_RC3_TEMPERATURE
0	9.74	59.69	85.08	22.41	22.10	23.19
10422	19.34	78.63	68.32	0.73	0.97	1.66
10603	15.28	63.40	64.88	1.46	0.88	1.77
11039	9.68	60.80	63.18	0.43	1.01	1.85
10734	18.47	0.00	62.91	1.03	0.99	1.86
11124	19.17	61.82	62.37	0.82	1.00	1.87
11326	8.31	69.56	61.98	0.56	0.53	1.57
11012	13.48	62.66	61.59	0.46	1.00	1.81
11409	15.66	72.13	61.52	0.61	1.02	2.02
10832	15.06	63.79	61.39	0.44	1.00	1.67
10981	13.79	61.52	61.33	0.88	0.88	1.71
11483	20.68	62.39	61.33	0.82	1.00	1.83
10749	13.75	72.33	61.27	0.62	1.07	1.74
11014	13.24	62.76	61.20	0.45	1.05	1.81
10690	14.49	62.25	61.10	0.48	0.91	1.84
11209	13.43	71.05	61.06	0.68	0.98	1.79
10733	17.02	68.70	60.97	0.81	0.90	1.79
10782	10.31	67.25	60.88	0.65	0.91	1.90
10153	14.24	62.66	60.75	0.98	0.96	1.73
10812	14.25	65.13	60.74	0.79	0.96	1.75
10356	18.53	66.72	60.61	0.91	0.98	1.77
10766	12.90	64.02	60.56	0.98	1.03	1.92
10275	15.10	74.26	60.42	0.61	0.93	1.43
10333	17.97	69.56	60.34	0.90	0.78	1.61
11037	13.13	63.35	60.30	0.60	1.00	1.87
10726	14.69	67.60	60.16	0.78	0.91	1.90
11219	15.76	64.41	60.14	0.58	1.01	1.78
Total	**18.95**	**57.06**	**62.06**	**13.03**	**12.85**	**13.56**

- On the basis of standard deviation derived in Power BI, python modules are written to find the best and the not ok coil.
- Values for RC1_Temperature shows least deviation while Linear_Potentiometer_Entry shows the highest deviation in values.

```
thisdata =pd.read_csv("SD_1month data.csv")
```

```
x = np.mean(thisdata)
print(x)
```

```
I_COIL_NO                                    28308.863174
Standard deviation of I_LINE_SPEED              12.569447
Standard deviation of I_LINEAR_POTENTIOMETER_ENTRY    43.186410
Standard deviation of I_LINEAR_POTENTIOMETER_EXIT     41.137337
Standard deviation of I_RC1_TEMPERATURE          0.661040
Standard deviation of I_RC2_TEMPERATURE          0.944204
Standard deviation of I_RC3_TEMPERATURE          1.738762
dtype: float64
```

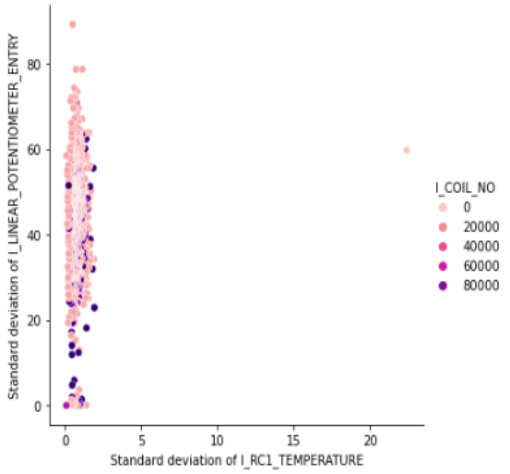
- When the data was sorted for each coil on the basis the standard deviation in each of the parameters in its descending order, **Coil no 0** shows the highest deviation while **Coil no 70239** shows least deviation for all the parameters.

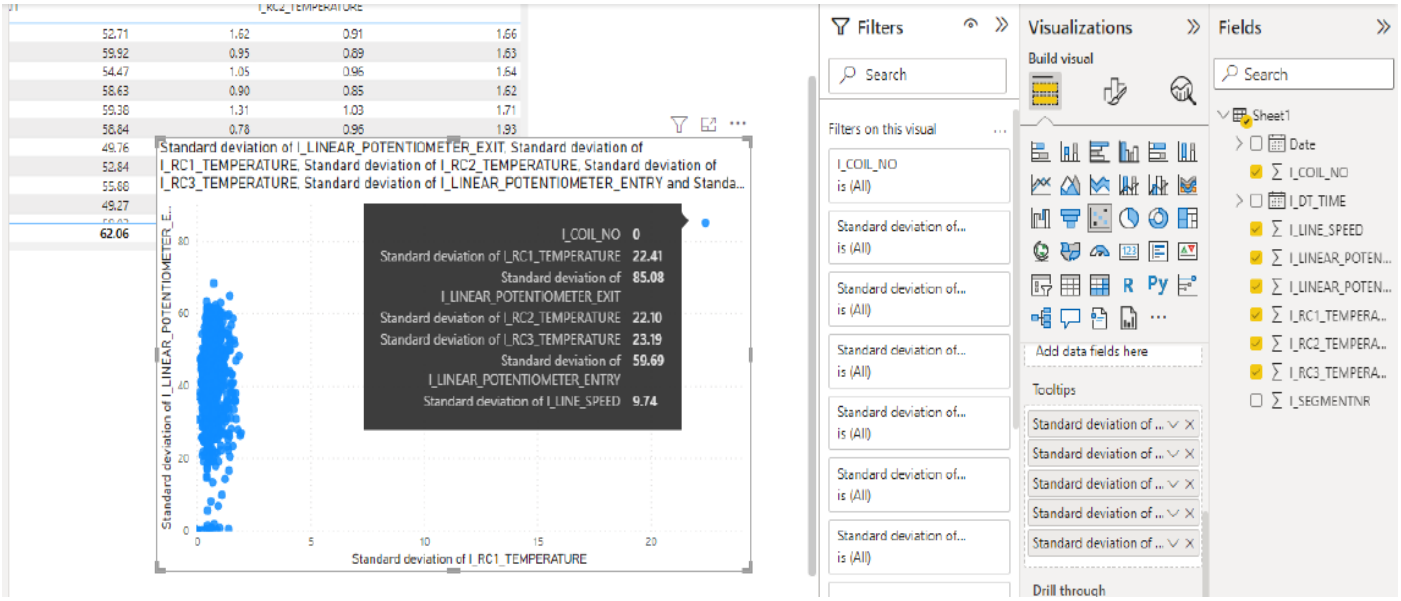
```
df= pd.DataFrame(thisdata)
df
```

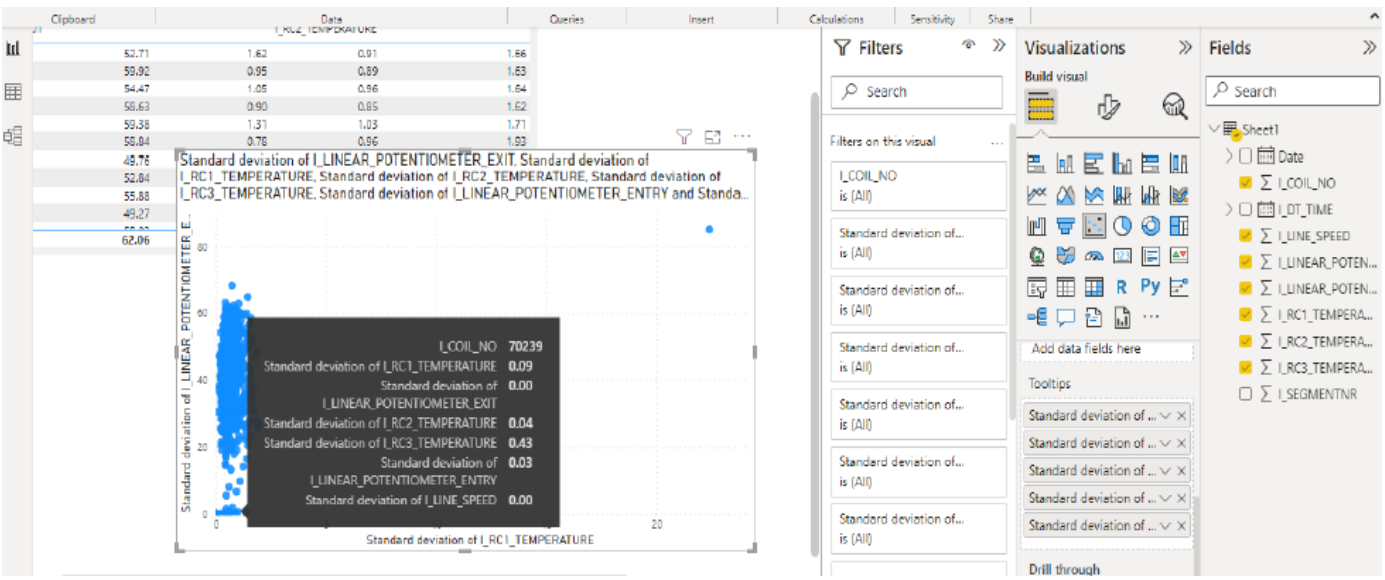
	I_COIL_NO	Standard deviation of I_LINE_SPEED	Standard deviation of I_LINEAR_POTENTIOMETER_ENTRY	Standard deviation of I_LINEAR_POTENTIOMETER_EXIT	Standard deviation of I_RC1_TEMPERATURE	Standard deviation of I_RC2_TEMPERATURE	Standar I_RC3_TE
0	0	9.739502	59.689178	85.077910	22.410254	22.102325	
1	99994	15.151346	22.902850	27.010655	1.933634	0.683927	
2	11460	14.919710	34.227966	26.141158	1.904704	0.999142	
3	99985	19.433759	55.557677	48.314519	1.863595	0.897090	
4	99024	15.690636	31.940996	27.837020	1.814277	0.960785	
...	...	...	...	...	...	...	
1274	11024	9.472202	53.984424	41.312039	0.183944	1.005697	
1275	11349	11.462232	42.481167	38.380395	0.159276	0.869757	
1276	10634	13.378999	55.006072	47.319056	0.155888	0.340656	
1277	11040	12.265640	58.424689	37.054083	0.098125	0.374772	
1278	70239	0.000000	0.026458	0.000000	0.091315	0.039609	

```
In [22]: sns.relplot(x = 'Standard deviation of I_RC1_TEMPERATURE' , y = 'Standard deviation of I_LINEAR_POTENTIOMETER_ENTRY' , hue = 'I_(
```

```
Out[22]: <seaborn.axisgrid.FacetGrid at 0x23374a3fa60>
```







Chapter 6

Result and Conclusion

The Data of 1 month Coil Data is analyzed i.e., total 1294 coils are analyzed on the basis of various Parameters. After cleaning the data by eliminating the null and duplicate values, 1290 coil were taken into consideration. From the data it is concluded that the Coil no 0 is found out to be not ok and coil number 70239 is found out to be best coil. Values for RC1_Temperature shows least deviation while Linear_Potentiometer_Entry shows the highest deviation in values.

References

1. H. Li, "Research on Big Data Analysis Data Acquisition and Data Analysis," 2021 International Conference on Artificial Intelligence, Big Data and Algorithms (CAIBDA), 2021, pp. 162-165, doi: 10.1109/CAIBDA53561.2021.00041.

2. J. Chen, Q. Jiang, Y. Wang and J. Tang, "Study of data analysis model based on big data technology," 2016 IEEE International Conference on Big Data Analysis (ICBDA), 2016, pp. 1-6, doi: 10.1109/ICBDA.2016.7509810.

3. X. HongJu, W. Fei, W. FenMei and W. XiuZhen, "Some key problems of data management in army data engineering based on big data," 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA), 2017, pp. 149-152, doi: 10.1109/ICBDA.2017.8078796

Acknowledgments

We owe our deepest gratitude and regards towards the ones who offered their valuable guidance in the hour of need. We thank our guide **Mr. Prashant Arote** (ITS Head) & **Mr. Rupesh Chandra** (Manager ITS) for his guidance and precious insights. His useful comments and feedbacks during the discussions we had and the encouragement to question every technical detail that we came across while completing this Internship project helped us to a great extent.