

# Data Analysis Report

[Data Link \( Working \)](#)

## 1. Introduction

This report presents a comprehensive analysis of a dataset containing various measurements. The objective is to explore the relationships between these measurements and derive insights using **descriptive statistics, visualizations, correlation analysis, data cleaning, pivot tables, and regression analysis.**

Index No	Measurement A	Measurement B	Measurement C	Measurement D	<u>Result</u>
1	5	365	8	60.5	<u>109.625</u>
2	25	355	10	63.5	<u>113.375</u>
3	6	360	12	66.5	<u>111.125</u>
4	13	365	14	69.5	<u>115.375</u>
5	13.5	370	16	72.5	<u>118</u>
6	14	375	18	75.5	<u>120.625</u>
7	14.5	380	20	78.5	<u>123.25</u>
8	15	385	22	81.5	<u>125.875</u>
9	15.5	390	24	84.5	<u>128.5</u>
10	16	395	26	87.5	<u>131.125</u>

## 2. Descriptive Statistics

### Overview

The descriptive statistics provide a summary of the dataset, including measures of central tendency and dispersion.

- **Mean:** Average value for each measurement.
- **Standard Error:** Accuracy of the mean estimate.
- **Median:** Middle value, showing the distribution's center.
- **Mode:** Most frequently occurring value.

- **Standard Deviation:** Dispersion from the mean.
- **Variance:** Measure of data spread.
- **Kurtosis & Skewness:** Shape of the data distribution.
- **Range:** Difference between maximum and minimum values.
- **Sum & Count:** Total and number of observations.
- **Confidence Level (95%):** Range in which the true mean likely falls.

## 2. Descriptive Statist

	Measurement A		Measurement B
	Mean	50.75483871	Mean
	Standard Error	8.434631674	Standard Error
	Median	20	Median
	Mode	20	Mode
	Standard Deviation	105.0103175	Standard Deviation
	Sample Variance	11027.16678	Sample Variance
	Kurtosis	7.266924563	Kurtosis
	Skewness	3.011026153	Skewness
	Range	420	Range
	Minimum	5	Minimum
	Maximum	425	Maximum
	Sum	7867	Sum
	Count	155	Count
	Largest(1)	425	Largest(1)
	Confidence Level(95.0%)	16.66251415	Confidence Level(95.0%)

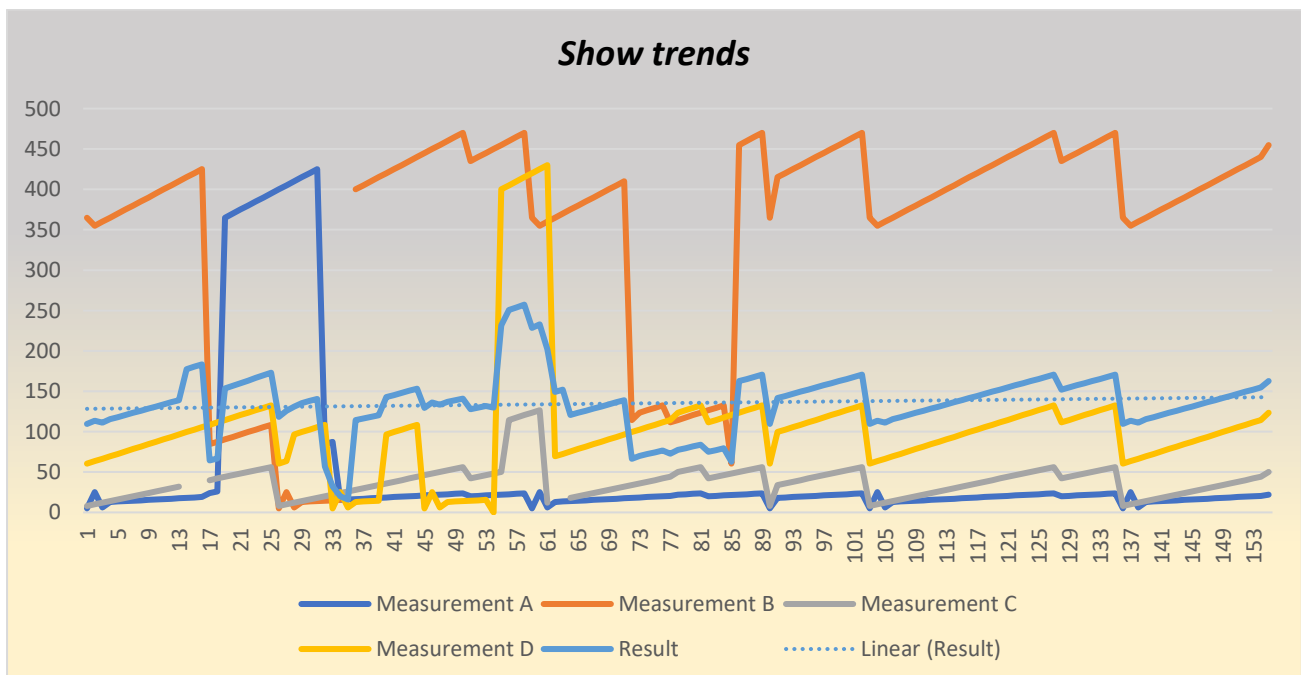
	Measurement B		Measurement C
Mean	346.4480519	Mean	37.72333333
Standard Error	11.33297679	Standard Error	1.711048037
Median	400	Median	38
Mode	365	Mode	42
Standard Deviation	140.6385434	Standard Deviation	20.95597308
Sample Variance	19779.1999	Sample Variance	439.1528076
Kurtosis	0.285486427	Kurtosis	6.099645153
Skewness	-1.357901296	Skewness	1.827123707
Range	465	Range	118.5
Minimum	5	Minimum	8
Maximum	470	Maximum	126.5
Sum	53353	Sum	5658.5
Count	154	Count	150
Largest(1)	470	Largest(1)	126.5
Confidence Level(95.0%)	22.38931919	Confidence Level(95.0%)	3.381053468

Measurement D		Result	
Mean	105.6233766	Mean	135.7674731
Standard Error	6.072099401	Standard Error	3.104141716
Median	101	Median	136.375
Mode	99.5	Mode	125.875
Standard Deviation	75.35277191	Standard Deviation	38.6462527
Sample Variance	5678.040234	Sample Variance	1493.532848
Kurtosis	10.81068823	Kurtosis	2.040065397
Skewness	3.001078503	Skewness	-0.008097927
Range	425	Range	241.25
Minimum	5	Minimum	16
Maximum	430	Maximum	257.25
Sum	16266	Sum	21043.95833
Count	154	Count	155
Largest(1)	430	Largest(1)	257.25
Confidence Level(95.0%)	11.99598077	Confidence Level(95.0%)	6.132194891

### 3. Visualization

Trend and Relationship Analysis

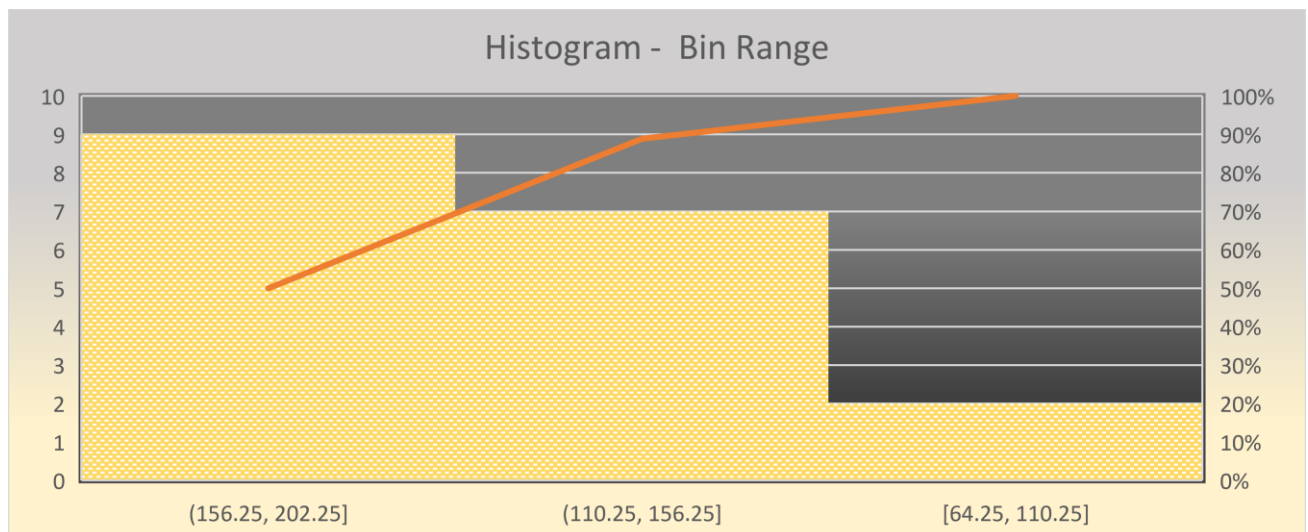
- **Line Chart:** Displays trends in measurements over time.
- **Scatter Plot:** Shows relationships between different measurements.
- **Histogram:** Visualizes the distribution of `Result` values, indicating frequency and distribution shape.





123.25	Frequency
64.25	6
66.75	2
125.875	46
128.5	5
131.125	7
133.75	6
136.375	6
139	7
153.5	26
156.75	5
160	7
163.25	6
166.5	5
169.75	5
173	5
177.5	2
180.3333333	1
183.1666667	1
More	7

In this Used a Data Analysis tool and Created a Histogram - Bin Range of the result.



## 4. Correlation Analysis

- **Correlation Coefficient (0.589)**: Indicates a moderate positive correlation between `Measurement A` and `Result`. This suggests that as `Measurement A` increases, `Result` tends to increase as well.

```
=CORREL(D2:D156,F2:F156)
```

### 4. Correlation Analysis

0.589014566

## 5. Data Cleaning

Data cleaning involved identifying and addressing missing values and outliers:

- **Missing Values**: Detected blank entries.
- **Outliers**: Identified using statistical measures, and removed to improve data quality.

5. Data Cleaning	( It found that Table is Blank)	
0	69.5	149.1666667
0	72.5	152
18	75.5	120.625
20	78.5	123.25
22	81.5	125.875
24	84.5	128.5
26	87.5	131.125
28	90.5	133.75

```
=IF(ISBLANK(0),"Missing",D63:F70)
```

## 6. Pivot Table

The pivot table provides a summary of the average results grouped by `Measurement A`, allowing for easy comparison of values.

6. Pivot Table	
Row Labels	Average of Result
24	64.25
26	66.75
84.5	56.875
87.5	32.375
<b>Grand Total</b>	<b>55.0625</b>

Value Filter (Measurement A) ? X

Show items for which

Average of Result  is between  25 and 85

OK Cancel

## 7. Regression Analysis

Regression Model Summary

- **Dependent Variable:** `Result`
- **Independent Variables:** Measurements
- **Multiple R (0.64):** Indicates a moderate linear relationship.
- **R Square (0.41):** 41% of the variability in `Result` is explained by the model.
- **ANOVA:** Shows the model's significance (p-value < 0.05).

ANOVA					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	1	360205.0293	360205.0293	106.0610216	3.18046E-19
Residual	153	519619.4481	3396.205543		
Total	154	879824.4774			

### Coefficients:

- **Intercept (-65.12):** Predicted `Result` when all measurements are zero.
- **Coefficient for `Result` (1.25):** For every one-unit increase in `Result`, the outcome increases by 1.25 units, holding other factors constant.

Dependent variable	Independent variables
SUMMARY OUTPUT	
Regression Statistics	
Multiple R	0.639848073
R Square	0.409405556
Adjusted R Square	0.405545462
Standard Error	58.27697267
Observations	155

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-65.1217442	17.16391059	-3.7941088	0.000212952	-99.03059973	-31.21288868	-99.03059973	-31.21288868
Result	1.255192074	0.121879955	10.29859318	3.18046E-19	1.014407215	1.495976933	1.014407215	1.495976933

## 8. Conclusions and Recommendations

- **Trends:** Positive trends between certain measurements and the result suggest potential predictive relationships.
- **Correlation:** Moderate correlations indicate some variables can be used as predictors.
- **Data Quality:** Addressing missing values and outliers improved analysis accuracy.
- **Model Validity:** The regression model provides a reasonable fit but could be improved with additional variables or refined techniques.