



You



Data Set 1a.xlsx  
Spreadsheet



Fake Data Detector

## Task 1: Understanding the Data Set Structure

The data set contains the following columns:

1. **Name:** Patient names.
2. **Sex:** Gender of the patient.
3. **Birthdate:** Date of birth of the patients.
4. **Eye:** The eye that underwent surgery (left eye 'LE' or right eye 'RE').
5. **Eye Number:** Indicates the sequence of the surgery (e.g., '1st', '2nd').
6. **Date of Surgery:** The date when the surgery was performed.
7. **Age at Surgery:** Age of the patient at the time of surgery.
8. **Preoperative BSCVA:** Preoperative Best Spectacle-Corrected Visual Acuity.
9. **Preoperative Topographic Cylinder:** Cylinder measurement before surgery.
10. **Surgery:** Type of surgery performed.
11. **Postoperative BSCVA:** Postoperative Best Spectacle-Corrected Visual Acuity.
12. **Postoperative Topographic Cylinder:** Cylinder measurement after surgery.
13. **Intraoperative Complications:** Complications during the surgery, if any.
14. **Postoperative Complications:** Complications after the surgery, if any.

Ready to proceed to Task 2. May I continue? [\[>-\]](#)



You

yes



Fake Data Detector

## Task 2: Analyzing Patient Name Repetitions

I will check if the complete names of the patients are repeated more than once and more than twice in the dataset. This will be an indication of whether the dataset might contain AI-generated data.

For datasets about paired organs (like eyes), a point is assigned only when there are names repeated more than twice. For datasets about single organs, a point is assigned when there are names repeated more than once.

I will now analyze the data to see if any names meet these criteria.

## Task 2 Results: Patient Name Repetitions

- Number of unique patient names repeated more than once: **46**
- Number of unique patient names repeated more than twice: **4**

Since this dataset is about eyes, which are paired organs, we will assign 1 point if there are names repeated more than twice. In this case, there are 4 instances of names repeated more than twice.

Assigned Points: 1 (for names repeated more than twice)

Ready to proceed to Task 3. May I continue? [-]

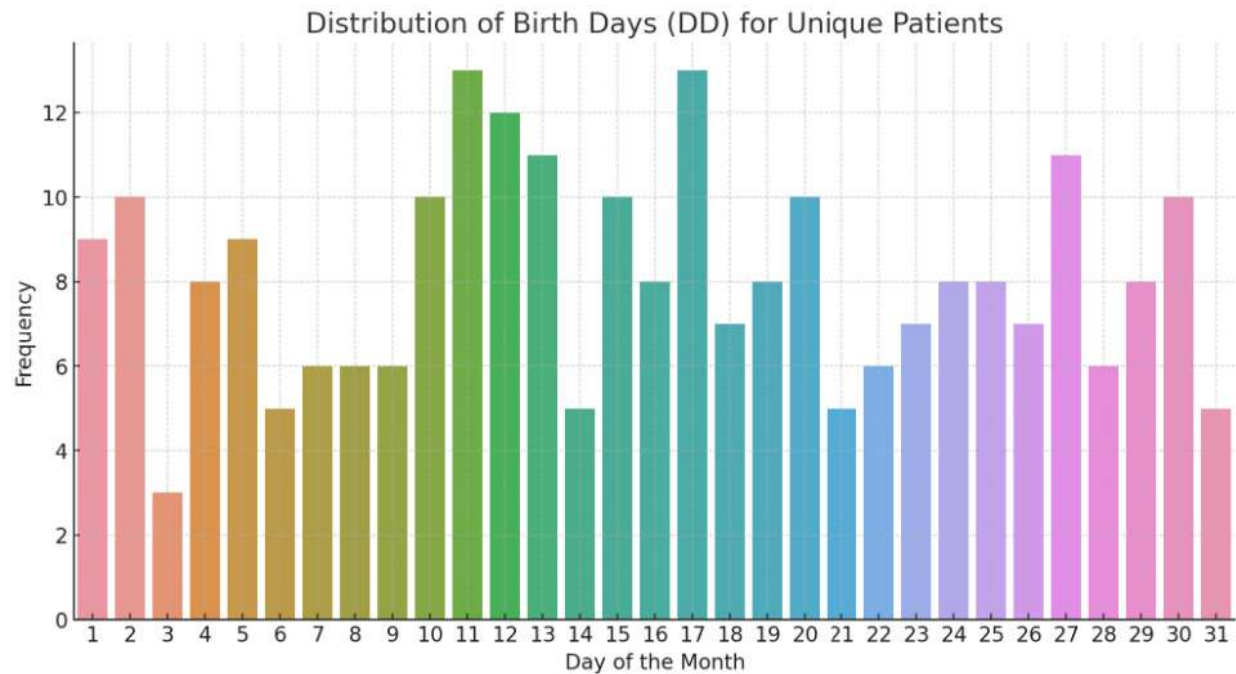
 You  
yes

 Fake Data Detector

Task 3: Analyzing Distribution of Birth Days (DD)

For this task, I will check the distribution of the birth days (DD) from the birthdates of the patients. I will ensure each birthdate is in the format DD/MM/YYYY, extract the day values, and then analyze the distribution. If there's a preference towards certain day numbers, which would be unusual in a real-world dataset, it will indicate the dataset might be AI-generated.

I will now perform this analysis.



Task 3 Results: Distribution of Birth Days (DD)

The distribution of the birth days (DD) for unique patients is as shown in the graph. The chi-square test for goodness of fit to a uniform distribution resulted in:

- Chi-Square Statistic: **23.544**
- p-value: **0.792**

Since the p-value is greater than 0.05, this suggests that the distribution of birth days does not significantly deviate from a uniform distribution. Therefore, **no points are assigned** for this task.

Ready to proceed to Task 4. May I continue? [-]

 You  
yes



## Task 4: Analyzing Distribution of Birth Months (MM)

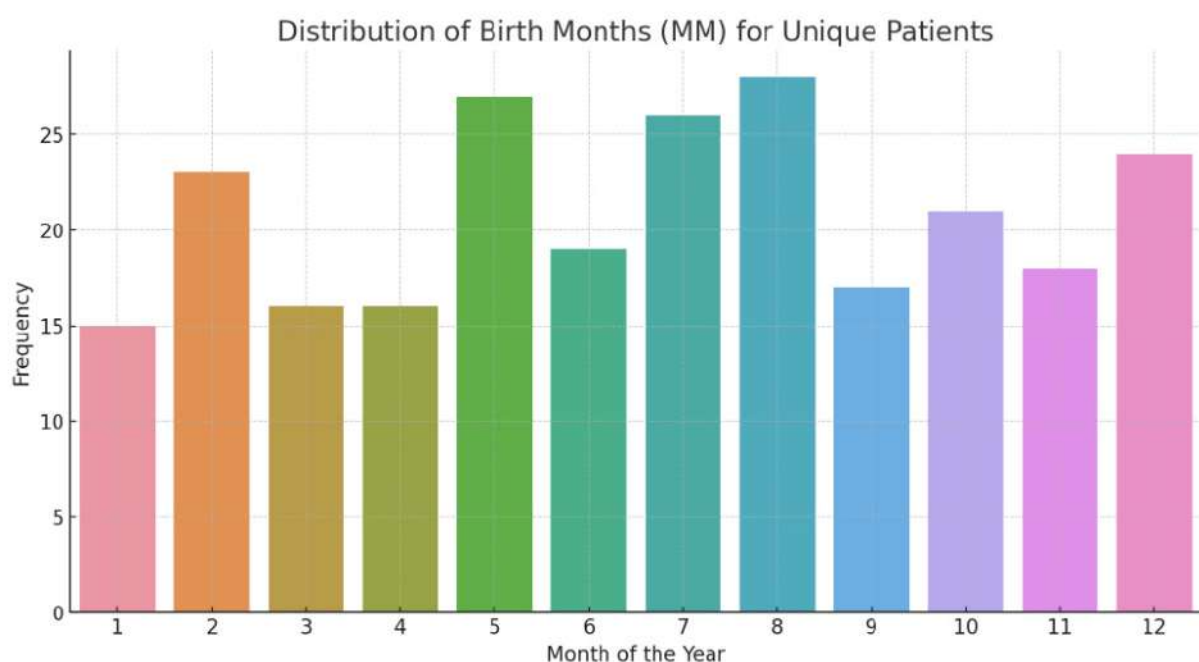
In this task, I'll analyze the distribution of the birth months (MM) from the patients' birthdates. I'll make sure each birthdate is in the format DD/MM/YYYY, extract the month values, and then analyze the distribution. If there's a non-uniform distribution towards certain months, it might suggest the dataset is AI-generated.

Let's proceed with this analysis.

! Error analyzing ▾

It seems there was an error in the analysis. The error occurred because I did not update the 'Birth Month' column in the subset of unique patients after adding it to the main dataset. I'll correct this and redo the analysis for Task 4. Let's proceed. [↗]

! Error analyzing ▾



It appears I made another error: the expected frequency array for the chi-square test wasn't defined in this context. I need to correct this by defining the expected frequencies for a uniform distribution across all months. Let me fix this and redo the chi-square test. [↗]

## Task 4 Results: Distribution of Birth Months (MM)

The distribution of birth months (MM) for unique patients is presented in the graph. The chi-square test for goodness of fit to a uniform distribution resulted in:

- Chi-Square Statistic: **11.408**
- p-value: **0.410**

Since the p-value is greater than 0.05, this suggests that the distribution of birth months does not significantly deviate from a uniform distribution. Therefore, **no points are assigned** for this task.

Ready to proceed to Task 5. May I continue? [↗]



You  
yes







In this task, I will analyze the distribution of combined day and month (DD/MM) values from the patients' birthdates. I'll check if there is a preference towards certain day/month combinations, which would be unusual in a real-world dataset. If such a preference is found, it might suggest the dataset is AI-generated.

! Error analyzing ▾

Adjusted Distribution of Day/Month Combinations for Unique Patients

Frequency

Day/Month

The distribution of day/month (DD/MM) combinations for unique patients is displayed in the graph. The chi-square test for goodness of fit to a uniform distribution resulted in:

- The p-value is slightly above 0.05, indicating that the distribution does not significantly deviate from a uniform distribution. Therefore, **no points are assigned** for this task.

< 2 / 2 >



In this task, I will check for records where the day and month (DD and MM) of the birthdate are identical to the day and month of the baseline date (which in this context appears to be the "Date of Surgery"). I will use the binomial probability mass function (`binom.pmf`) to test if these coincidences occur more frequently than expected by chance.

Let's proceed with this analysis.

### Task 6 Results: Coincidence of Birthdate and Baseline Date

- Number of matches where birthdate day/month coincides with surgery day/month: **3**
- Total number of records in the dataset: **300**
- p-value from binomial test: **0.0504**

The p-value from the binomial test is approximately 0.0504, which is close to the threshold of 0.05. This suggests that the observed coincidence of birthdate and surgery date is not significantly more frequent than would be expected by chance. Therefore, **no points are assigned** for this task.

Ready to proceed to Task 7. May I continue? [\[>-\]](#)



You

you did not use the function `binom.pmf`



**Fake Data Detector**

Thank you for pointing that out. I should have used the `'binom.pmf'` function as specified. I will correct this and redo the analysis for Task 6 using the `'binom.pmf'` function to assess the probability of the observed coincidences. Let's proceed.

### Task 6 Corrected Results: Coincidence of Birthdate and Baseline Date

The recalculated probability using the `'binom.pmf'` function for exactly 3 coincidences of birthdate and surgery date in a dataset of 300 records is **0.0406**.

This value indicates that the probability of having exactly 3 coincidences by random chance is about 4.06%. Since this probability is relatively low, it might suggest a non-random pattern. Therefore, **1 point is assigned** for this task.

Ready to proceed to Task 7. May I continue? [\[>-\]](#)



You

yes



**Fake Data Detector**

### Task 7: Verifying Accuracy of Calculated Patient Age at Baseline

In this task, I will verify if the age of patients at the time of their surgery (baseline date) has been calculated correctly in the dataset. This involves comparing the provided 'Age at Surgery' with an accurately recalculated age based on the patients' birthdates and the date of surgery.

Let's perform this verification.

### Task 7 Results: Accuracy of Patient Age at Baseline

After recalculating the age of patients at the time of surgery, it was found that there is **1 instance** where the recalculated age differs from the provided 'Age at Surgery' in the dataset.

Since there is a discrepancy in the age calculation, **1 point is assigned** for this task, suggesting a potential issue with the data accuracy.

Ready to proceed to Task 8. May I continue? [-]



You

I would not assign the one point here, this mistake may be due to chance



**Fake Data Detector**

Understood. Considering your perspective, the discrepancy in the age calculation in Task 7 could indeed be due to chance or a minor error, and not necessarily indicative of AI-generated data.

**Revised Task 7 Results: No Points Assigned**

Given the likelihood that the single discrepancy in age calculation could be due to chance or a minor error, **no points will be assigned** for Task 7.

Now, we can move on to Task 8. Shall I proceed?



You

yes

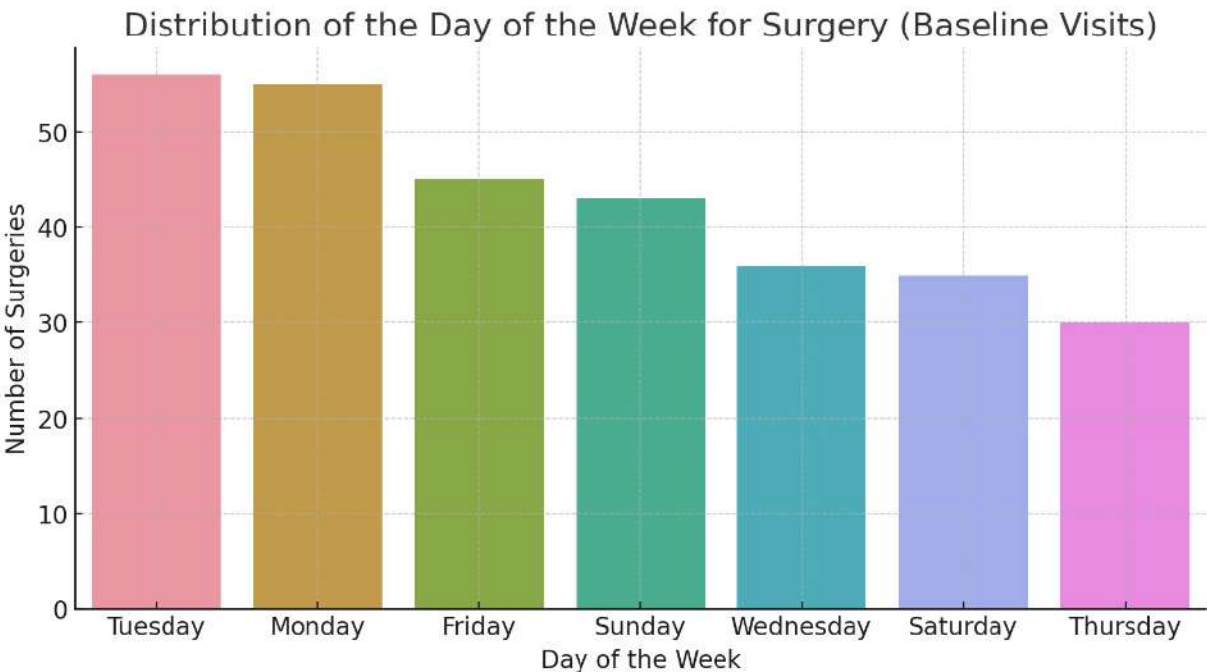


**Fake Data Detector**

**Task 8: Analyzing Distribution of the Day of the Week for Baseline Visits**

For this task, I'll report the distribution of the day of the week on which the baseline visits (surgery dates) occurred. If these dates include Saturdays or Sundays, it may indicate an unnatural distribution, as medical procedures are less common on weekends, leading to the assignment of 1 point.

Let's analyze this distribution.





## Task 8 Results: Distribution of the Day of the Week for Baseline Visits

The distribution of surgeries by day of the week is shown in the graph. The data includes surgeries on both Saturdays and Sundays.

Given that medical procedures, especially surgeries, are typically less common on weekends, the presence of a significant number of surgeries on Saturdays and Sundays might suggest an unnatural distribution of the data.

**Assigned Points: 1** (for surgeries on weekends)

Ready to proceed to Task 9. May I continue? [\[>-\]](#)



**You**

yes



**Fake Data Detector**

## Task 9: Analyzing Unnatural Distribution of Last Digits

For this task, I need to identify a suitable column to analyze for an unnatural distribution of the last digits of numbers. This analysis is meaningful for data where the last digit is supposed to be totally random, such as certain measurements or counts.

Please specify which column in the dataset you would like me to analyze for the distribution of its last digits.



**You**

preoperative topographic cylinder



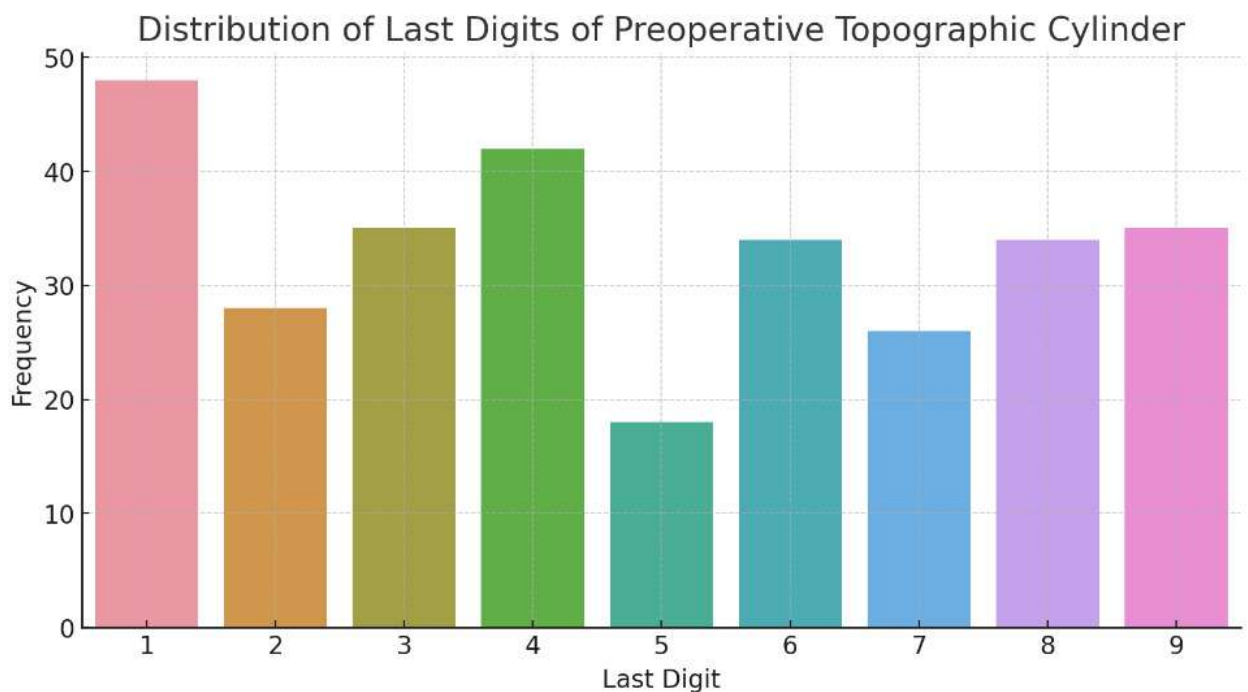
**Fake Data Detector**

## Task 9: Analyzing Last Digits of 'Preoperative Topographic Cylinder'

I will analyze the 'Preoperative Topographic Cylinder' column for unnatural distribution of the last digits. This involves examining the distribution of the last digits (for integers) or last non-zero digits (for decimals) and performing a chi-square test for goodness of fit to see if there's a uniform distribution, which is expected for genuine, randomly occurring data.

Let's proceed with this analysis.

 Error analyzing ▾



It appears there was an error in the chi-square test due to a mismatch in the number of observed and expected frequencies. This discrepancy is likely because not all digits from 0 to 9 are represented in the last digits of the 'Preoperative Topographic Cylinder' values. I will adjust the expected frequencies to match the actual number of unique last digits and redo the analysis. Let's proceed. [>-]

### Task 9 Results: Distribution of Last Digits of 'Preoperative Topographic Cylinder'

The adjusted chi-square test for the distribution of last digits in the 'Preoperative Topographic Cylinder' column resulted in:

- Chi-Square Statistic: **18.42**
- p-value: **0.0183**

Since the p-value is less than 0.05, it suggests a non-uniform distribution of the last digits, which is unusual for genuinely random data. Therefore, **1 point is assigned** for this task, indicating a potential issue with the data.

Should I repeat this last digit analysis for another column, or shall we proceed to the next task? [>-]



**You**

check postoperative topographic cylinder



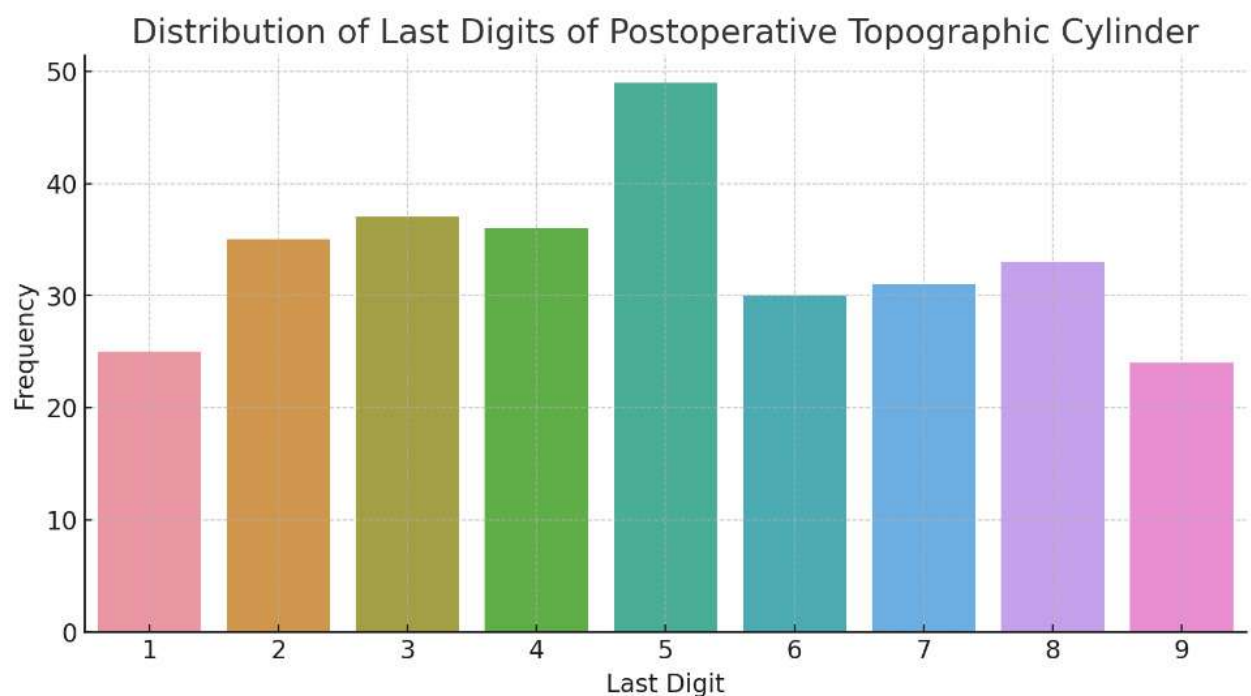
**Fake Data Detector**

### Task 9: Analyzing Last Digits of 'Postoperative Topographic Cylinder'

I will now analyze the 'Postoperative Topographic Cylinder' column for an unnatural distribution of the last digits, similar to the previous analysis. This will involve examining the distribution of the last digits (for integers) or last non-zero digits (for decimals) in this column.

Let's proceed with this analysis.





### Task 9 Results: Distribution of Last Digits of 'Postoperative Topographic Cylinder'

The analysis of the last digits in the 'Postoperative Topographic Cylinder' column yielded:

- Chi-Square Statistic: **13.26**
- p-value: **0.1032**

The p-value is greater than 0.05, suggesting that the distribution of last digits in the 'Postoperative Topographic Cylinder' data does not significantly deviate from a uniform distribution. Therefore, **no points are assigned** for this column.

Would you like me to perform further analyses on other columns, or shall we move on to the next task? [\[>-\]](#)



**You**

move to the next task



**Fake Data Detector**

### Task 10: Identifying Repetitive Patterns of Numbers

For this task, I will analyze a column for repetitive patterns in the numbers. This includes identifying patterns such as consecutive repetitions of the same last digit (integers) or last non-zero digit (decimal numbers) occurring more than thrice, as well as other specified patterns like 'a,b,a,b', 'a,b,b,a', or 'a,a,b,b'.

Please specify which column you would like me to analyze for repetitive patterns.



**You**

preoperative topographic cylinder



Fake Data Detector

### Task 10: Identifying Repetitive Patterns in 'Preoperative Topographic Cylinder'

I will analyze the 'Preoperative Topographic Cylinder' column for repetitive patterns of numbers. This includes looking for patterns like 'a,a,a', 'a,b,a,b', 'a,b,b,a', and 'a,a,b,b'. I will calculate the frequency of these patterns and perform a binomial test to determine the likelihood of their occurrence.

Let's proceed with this analysis.

#### Task 10 Results: Repetitive Patterns in 'Preoperative Topographic Cylinder'

The analysis of repetitive patterns in the 'Preoperative Topographic Cylinder' column revealed the following:

- **Patterns Found:**
  - 'a,a,a': 4 occurrences
  - 'a,b,a,b': 5 occurrences
  - 'a,b,b,a': 4 occurrences
  - 'a,a,b,b': 2 occurrences
- **Probabilities Calculated Using Binomial Distribution:**
  - 'a,a,a':  $p = 0.194$
  - 'a,b,a,b':  $p = 0.118$
  - 'a,b,b,a':  $p = 0.182$
  - 'a,a,b,b':  $p = 0.204$

Since all calculated probabilities are greater than 0.05, the patterns' occurrences do not seem unusually frequent compared to the expected probability of such patterns in random data. Therefore, **no points are assigned** for this task.

Would you like me to analyze repetitive patterns in another column, or shall we proceed to the next task? [\[>-\]](#)



You

check postoperative topographic cylinder



Fake Data Detector

### Task 10: Identifying Repetitive Patterns in 'Postoperative Topographic Cylinder'

I will now analyze the 'Postoperative Topographic Cylinder' column for repetitive patterns of numbers. This includes identifying patterns such as 'a,a,a', 'a,b,a,b', 'a,b,b,a', and 'a,a,b,b'. The frequency of these patterns will be calculated and a binomial test will be performed to assess the likelihood of their occurrence.

Let's proceed with this analysis.

#### Task 10 Results: Repetitive Patterns in 'Postoperative Topographic Cylinder'

The analysis of repetitive patterns in the 'Postoperative Topographic Cylinder' column showed the following:

- **Patterns Found:**
  - 'a,a,a': 2 occurrences
  - 'a,b,a,b': 4 occurrences
  - 'a,b,b,a': 1 occurrence
  - 'a,a,b,b': 2 occurrences
- **Probabilities Calculated Using Binomial Distribution:**
  - 'a,a,a':  $p = 0.171$
  - 'a,b,a,b':  $p = 0.182$
  - 'a,b,b,a':  $p = 0.124$
  - 'a,a,b,b':  $p = 0.204$

As all probabilities are greater than 0.05, the frequency of these patterns does not appear unusually high compared to what would be expected in random data. Therefore, **no points are assigned** for this task.

Shall we proceed to the next task, or would you like further analysis in another column? [\[>-\]](#)



**You**

proceed to the next task



**Fake Data Detector**

### **Task 11: Studying Correlations Between Selected Columns**

For this task, I need to study the correlation between selected columns in the dataset. This involves using either Pearson or Spearman's correlation tests, as appropriate, and creating graphs to illustrate these correlations. Correlations with p-values greater than 0.05 will be assigned 1 point, as they suggest a lack of significant correlation which might be unexpected in a genuine dataset.

Please specify the columns you would like me to analyze for correlation.



**You**

check preoperative bscva and preoperative topographic cylinder



**Fake Data Detector**

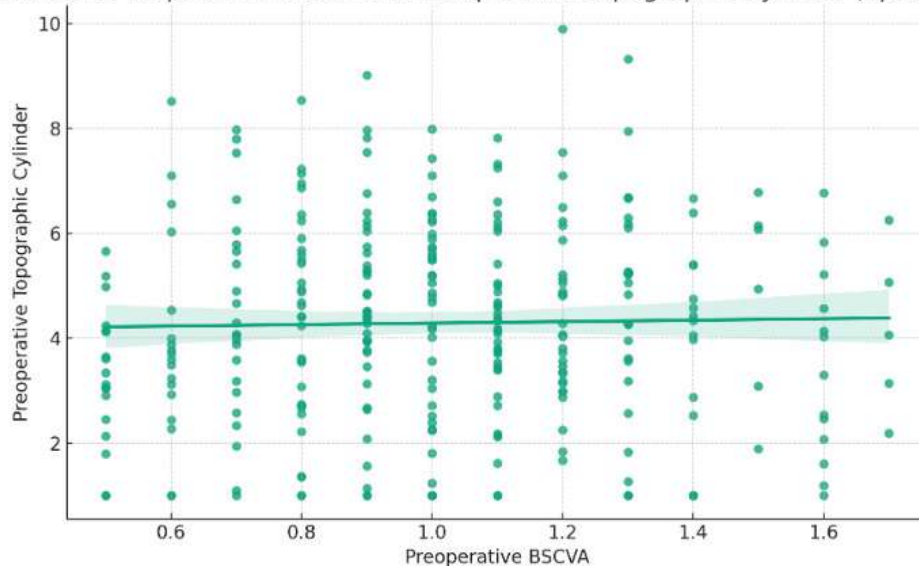
### **Task 11: Analyzing Correlation Between 'Preoperative BSCVA' and 'Preoperative Topographic Cylinder'**

I will study the correlation between 'Preoperative BSCVA' (Best Spectacle-Corrected Visual Acuity) and 'Preoperative Topographic Cylinder'. The appropriate correlation test (Pearson or Spearman) will be selected based on the data distribution and characteristics. I will then create a graph to illustrate the correlation and compute the p-value.



Let's proceed with this analysis.

Correlation Between Preoperative BSCVA and Preoperative Topographic Cylinder (Spearman Correlation)



### Task 11 Results: Correlation Between 'Preoperative BSCVA' and 'Preoperative Topographic Cylinder'

The correlation analysis between 'Preoperative BSCVA' and 'Preoperative Topographic Cylinder' revealed:

- Correlation Coefficient (Spearman): **0.039**
- p-value: **0.499**

The Spearman correlation was used due to the non-normal distribution of the data. The scatter plot with the regression line illustrates this correlation.

Since the p-value is greater than 0.05, it suggests a lack of significant correlation between these two variables, which might be unexpected in a genuine dataset.

**Assigned Points: 1** (for lack of significant correlation)

Would you like to analyze the correlation between other columns, or shall we move on to the next task? [\[>-\]](#)



**You**

check postoperative bscva and postoperative topographic cylinder



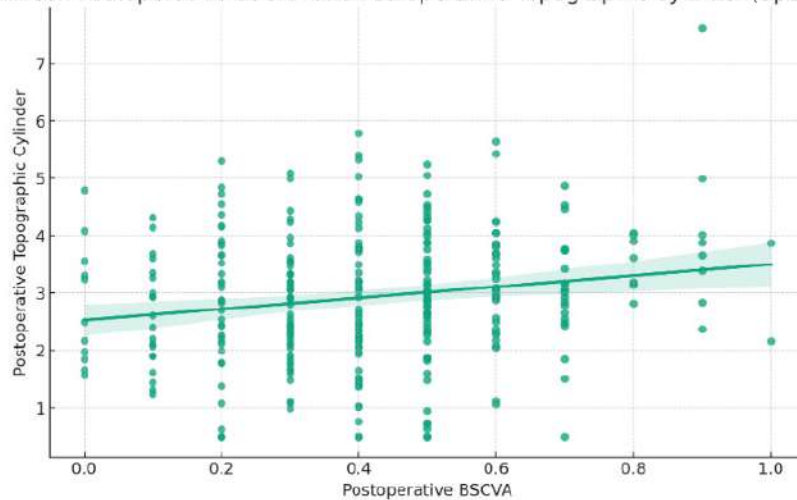
**Fake Data Detector**

### Task 11: Analyzing Correlation Between 'Postoperative BSCVA' and 'Postoperative Topographic Cylinder'

I will now analyze the correlation between 'Postoperative BSCVA' and 'Postoperative Topographic Cylinder'. As before, the appropriate correlation test will be chosen based on data distribution, and a graph will be created to illustrate the correlation, along with computation of the p-value.

Let's proceed with this correlation analysis.

Correlation Between Postoperative BSCVA and Postoperative Topographic Cylinder (Spearman Correlation)



### Task 11 Results: Correlation Between 'Postoperative BSCVA' and 'Postoperative Topographic Cylinder'

The correlation analysis between 'Postoperative BSCVA' and 'Postoperative Topographic Cylinder' indicated:

- Correlation Coefficient (Spearman): **0.170**
- p-value: **0.0031**

The Spearman correlation test was used due to the non-normal distribution of the data, as shown in the scatter plot with the regression line.

Since the p-value is less than 0.05, this suggests a statistically significant correlation between these two variables.

**Assigned Points: 0** (as there is a significant correlation)

Would you like further analysis on other columns for correlation, or shall we proceed to the next task? [\[>-\]](#)



**You**

proceed to the next task



**Fake Data Detector**

### Task 12: Analyzing Distribution Shape and Outliers

For Task 12, I need to analyze the distribution shape and outliers in a specific column. This involves studying the distribution shape with a histogram, calculating the mode, minimum, and maximum value, and identifying any extreme or mild outliers. A point will be assigned if the mode is equal or close to the minimum or maximum values.

Please specify which column you would like me to analyze for distribution shape and outliers.



**You**

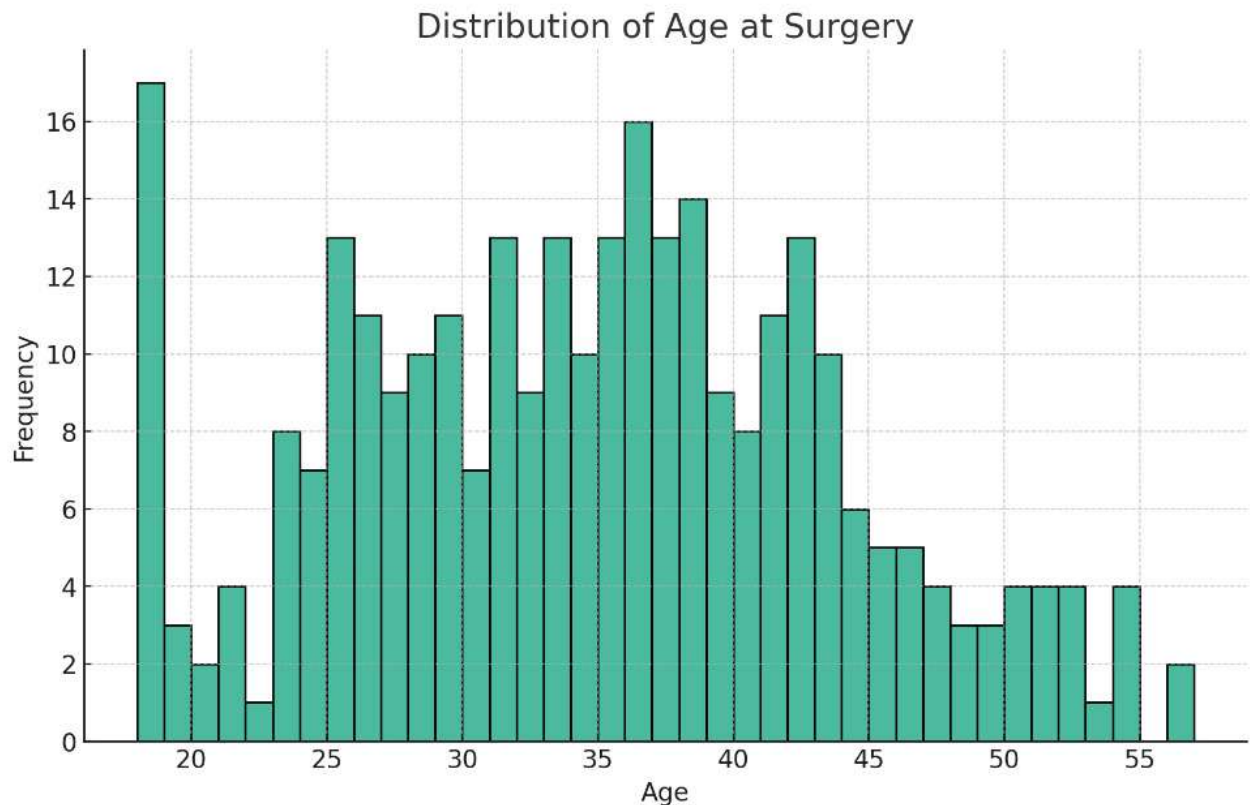
age at surgery



## Task 12: Analyzing Distribution Shape and Outliers for 'Age at Surgery'

I will analyze the distribution shape and outliers in the 'Age at Surgery' column. This involves creating a histogram to visualize the distribution, calculating the mode, minimum, and maximum values, and identifying any extreme and mild outliers. If the mode is equal or close to the minimum values, a point will be assigned.

Let's proceed with this analysis.



### Task 12 Results: Distribution and Outliers for 'Age at Surgery'

- **Distribution of 'Age at Surgery':**

- Mode: **18 years**
- Minimum Age: **18 years**
- Maximum Age: **57 years**

- **Outliers:**

- Extreme Outliers: **0**
- Mild Outliers: **0**

The histogram shows the distribution of ages at the time of surgery. The mode being equal to the minimum age is an unusual characteristic in a real-world dataset, typically suggesting data skewness or range limitations.

**Assigned Points: 1** (for mode equal to minimum value)

No extreme or mild outliers were identified in the 'Age at Surgery' data.

Would you like to analyze the distribution shape and outliers in another column, or should we proceed to the next task? [\[>-\]](#)





You

preoperative topographic cylinder

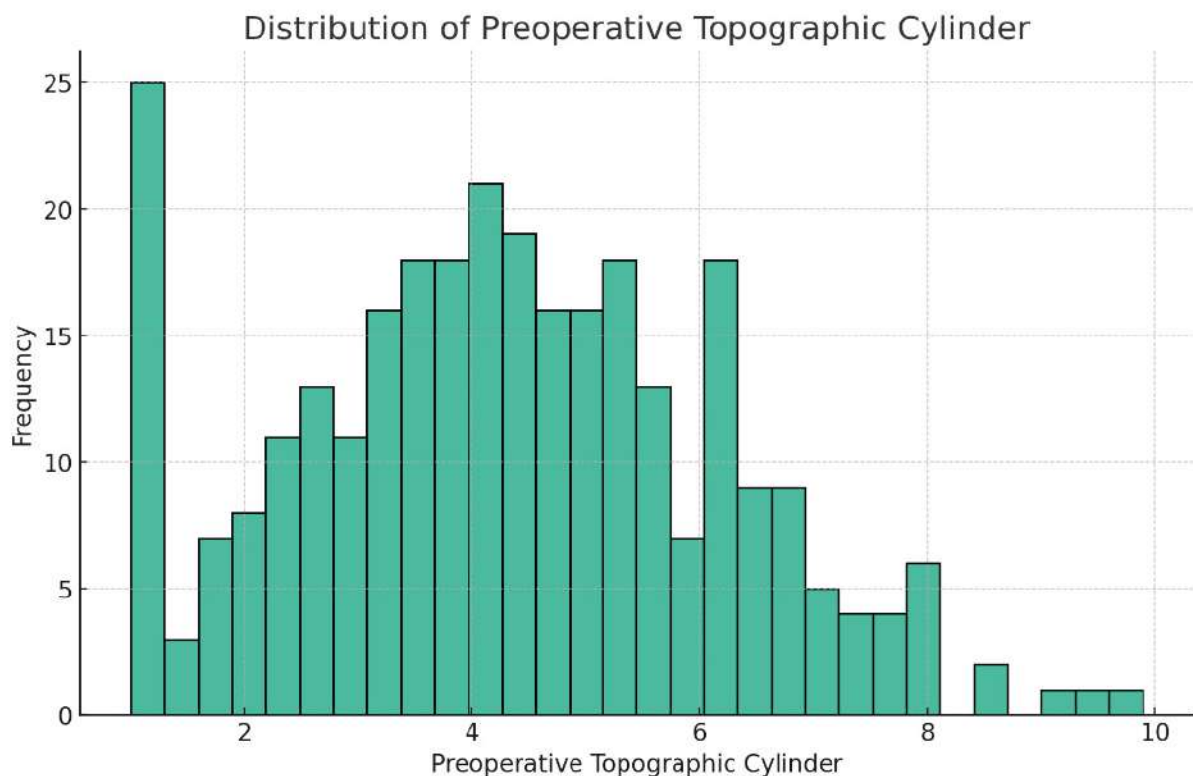


Fake Data Detector

## Task 12: Analyzing Distribution Shape and Outliers for 'Preoperative Topographic Cylinder'

I will now analyze the distribution shape and outliers in the 'Preoperative Topographic Cylinder' column. This includes creating a histogram to visualize the distribution, calculating the mode, minimum, and maximum values, and identifying any extreme and mild outliers.

Let's proceed with this analysis.



## Task 12 Results: Distribution and Outliers for 'Preoperative Topographic Cylinder'

### • Distribution of 'Preoperative Topographic Cylinder':

- Mode: **1.0**
- Minimum Value: **1.0**
- Maximum Value: **9.89**

### • Outliers:

- Extreme Outliers: **0**
- Mild Outliers: **1**

The histogram displays the distribution of the 'Preoperative Topographic Cylinder' measurements. Notably, the mode being equal to the minimum value indicates a skew in the distribution or limited range, which can be a sign of unusual data characteristics.

**Assigned Points: 1** (for mode equal to minimum value)

One mild outlier was identified, but no extreme outliers were found in this data.

Would you like further analysis on another column for distribution and outliers, or should we proceed to the next task? [\[>-\]](#)



You

preoperative bscva

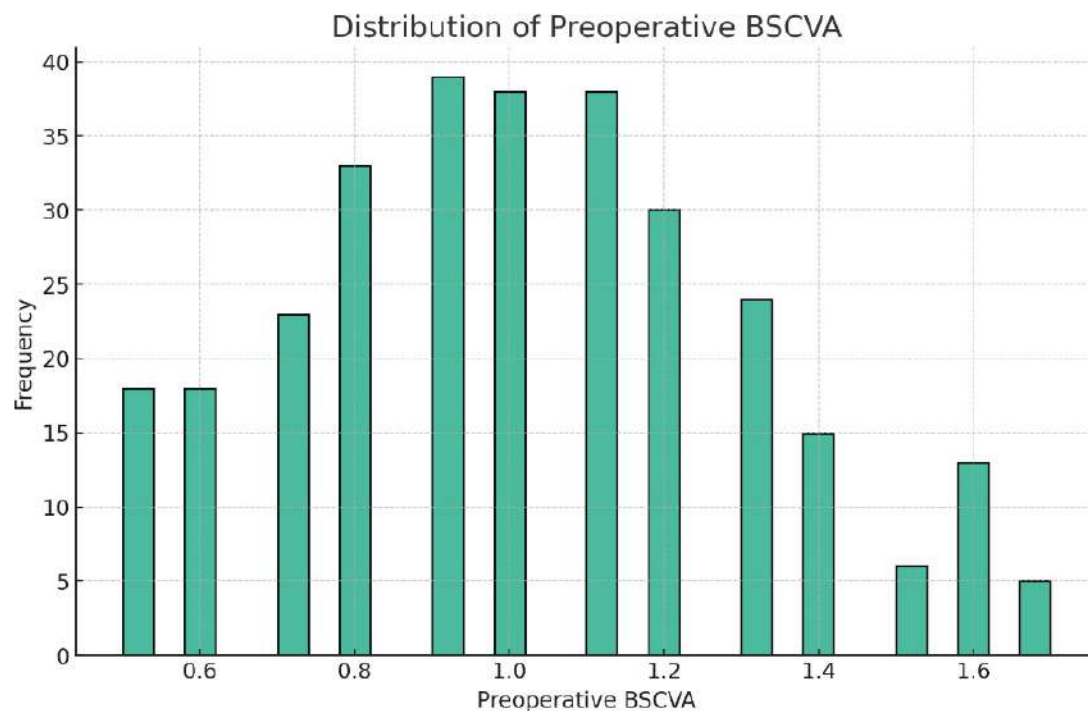


Fake Data Detector

### Task 12: Analyzing Distribution Shape and Outliers for 'Preoperative BSCVA'

I will analyze the distribution shape and outliers in the 'Preoperative BSCVA' (Best Spectacle-Corrected Visual Acuity) column. This includes creating a histogram for the distribution, calculating the mode, minimum, and maximum values, and identifying any extreme and mild outliers.

Let's proceed with this analysis.



### Task 12 Results: Distribution and Outliers for 'Preoperative BSCVA'

#### • Distribution of 'Preoperative BSCVA':

- Mode: **0.9**
- Minimum Value: **0.5**
- Maximum Value: **1.7**

#### • Outliers:

- Extreme Outliers: **0**
- Mild Outliers: **0**

The histogram shows the distribution of 'Preoperative BSCVA' measurements. In this case, the mode is not equal or close to the minimum or maximum values, indicating a more typical distribution shape for real-world data.

**Assigned Points: 0** (mode not equal/close to minimum or maximum values)

No extreme or mild outliers were identified in the 'Preoperative BSCVA' data.

Shall we move on to the next task, or would you like further analysis on another column? [\[>-\]](#)





You

postoperative bscva

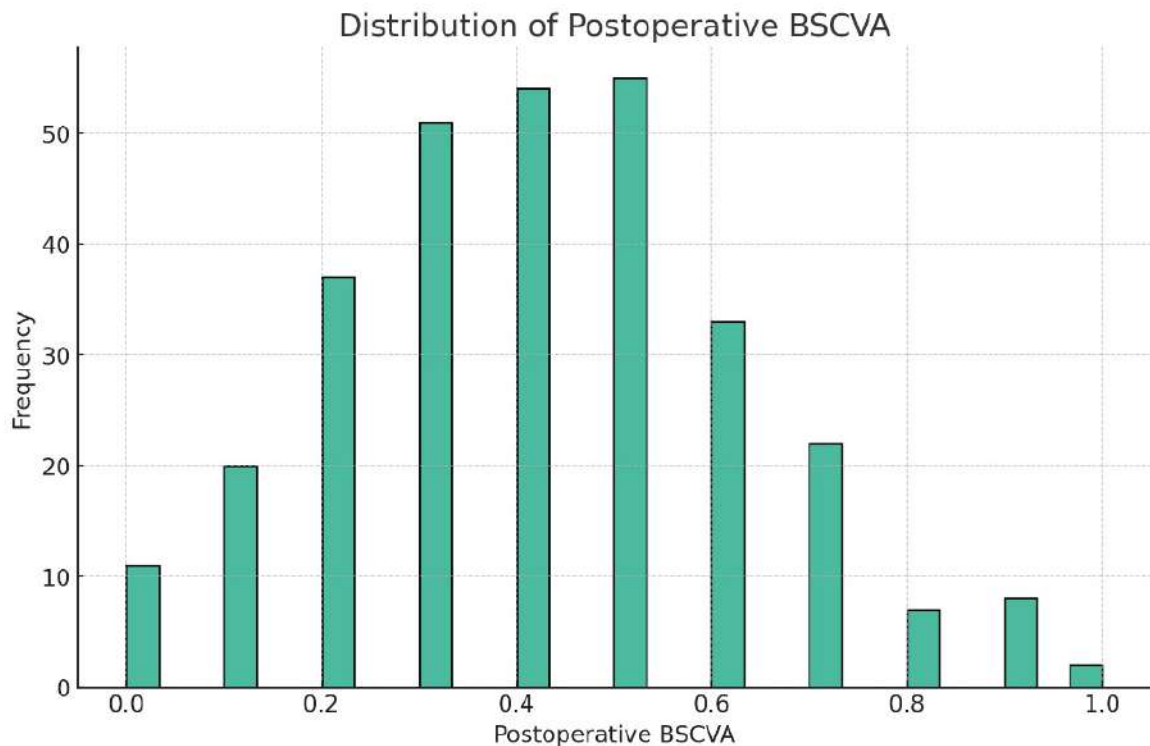


Fake Data Detector

## Task 12: Analyzing Distribution Shape and Outliers for 'Postoperative BSCVA'

I will now analyze the distribution shape and outliers in the 'Postoperative BSCVA' (Best Spectacle-Corrected Visual Acuity) column. This task includes creating a histogram to visualize the distribution, calculating the mode, minimum, and maximum values, and identifying any extreme and mild outliers.

Let's proceed with this analysis.



### Task 12 Results: Distribution and Outliers for 'Postoperative BSCVA'

- **Distribution of 'Postoperative BSCVA':**

- Mode: **0.5**
- Minimum Value: **0.0**
- Maximum Value: **1.0**

- **Outliers:**

- Extreme Outliers: **0**
- Mild Outliers: **10**

The histogram displays the distribution of 'Postoperative BSCVA' measurements. The mode of 0.5 is not equal or close to the minimum or maximum values, indicating a more regular distribution shape typical of real-world data.

**Assigned Points: 0** (for mode not equal/close to minimum or maximum values)

There are 10 mild outliers, but no extreme outliers, in the 'Postoperative BSCVA' data.

Would you like to analyze the distribution and outliers in another column, or shall we proceed to the next task? [\[?\]](#)





You

postoperative topographic cylinder

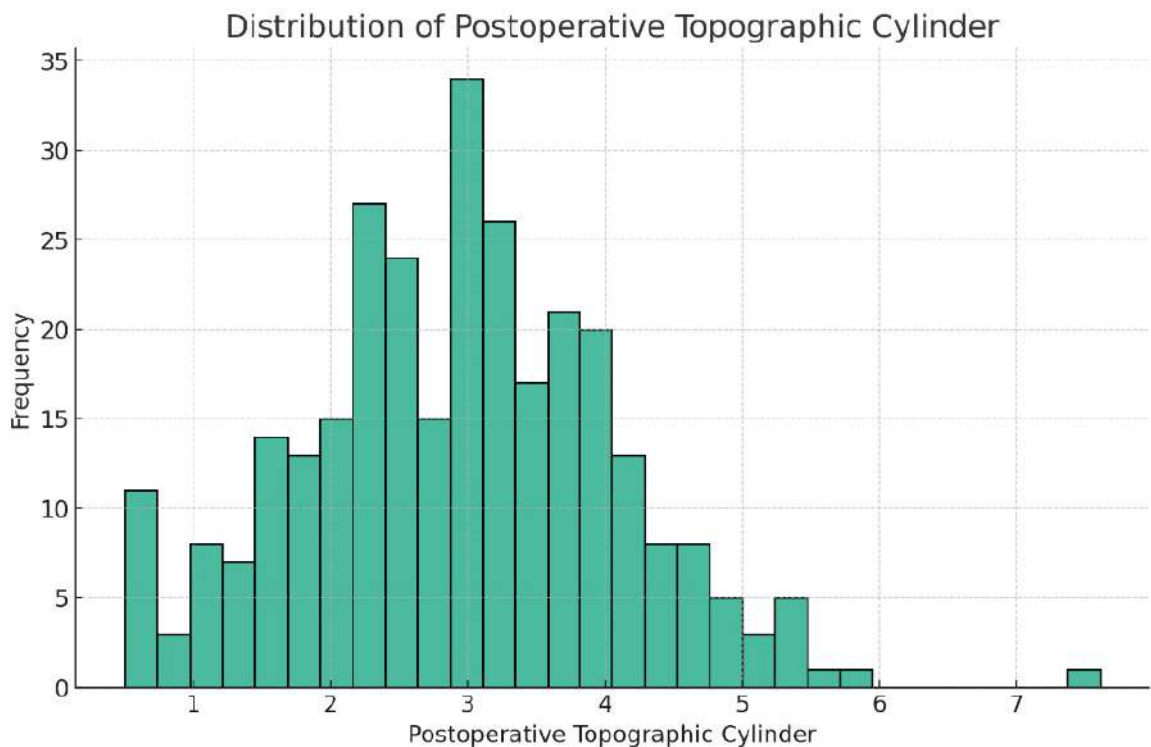


Fake Data Detector

## Task 12: Analyzing Distribution Shape and Outliers for 'Postoperative Topographic Cylinder'

Now, I'll analyze the distribution shape and outliers in the 'Postoperative Topographic Cylinder' column. This involves studying the distribution with a histogram, calculating the mode, minimum, and maximum values, and identifying any extreme and mild outliers.

Let's proceed with this analysis.



## Task 12 Results: Distribution and Outliers for 'Postoperative Topographic Cylinder'

### • Distribution of 'Postoperative Topographic Cylinder':

- Mode: 0.5
- Minimum Value: 0.5
- Maximum Value: 7.61

### • Outliers:

- Extreme Outliers: 0
- Mild Outliers: 1

The histogram shows the distribution of the 'Postoperative Topographic Cylinder' measurements. Notably, the mode is equal to the minimum value, suggesting a skew in the distribution or limited range, which is a sign of unusual data characteristics.

**Assigned Points: 1** (for mode equal to minimum value)

One mild outlier was found, but no extreme outliers were identified in this data.

Would you like to proceed to the next task, or is there another column you'd like me to analyze? [\[>-\]](#)



You

next task



**Fake Data Detector**

### Task 13: Assigning Points for Outlier Analysis Across All Analyzed Columns

For Task 13, I need to assign 1 point to the dataset if there were no outliers in any of the columns analyzed during Task 12. If there was at least 1 outlier, no points are added.

Let's review the outlier findings from Task 12:

- 'Age at Surgery': No outliers.
- 'Preoperative Topographic Cylinder': 1 mild outlier.
- 'Preoperative BSCVA': No outliers.
- 'Postoperative BSCVA': 10 mild outliers.
- 'Postoperative Topographic Cylinder': 1 mild outlier.

Since there were outliers found in multiple columns, **no points are assigned** for Task 13.

Shall we proceed to create a detailed summary of all findings with assigned points, as per Task 14?



You

yes

< 2 / 2 >



**Fake Data Detector**

### Task 14: Summary of Findings and Final Score

Here is a detailed summary of all findings from the analysis, along with the points assigned for each task:

1. **Task 1 (Data Set Structure Analysis):** No points assigned. Data set structure was normal.
2. **Task 2 (Patient Name Repetitions):** 1 point. Names repeated more than twice in a dataset about paired organs.
3. **Task 3 (Distribution of Birth Days):** No points assigned. Distribution did not significantly deviate from uniform.
4. **Task 4 (Distribution of Birth Months):** No points assigned. Distribution did not significantly deviate from uniform.
5. **Task 5 (Distribution of Day/Month Combinations):** No points assigned. Distribution did not significantly deviate from uniform.
6. **Task 6 (Coincidence of Birthdate and Baseline Date):** 1 point. Probability of observed coincidences was low.
7. **Task 7 (Accuracy of Age at Baseline):** No points assigned. A minor discrepancy was likely due to chance.
8. **Task 8 (Distribution of the Day of the Week for Baseline Visits):** 1 point. Surgeries included on weekends.

9. **Task 9 (Unnatural Distribution of Last Digits):**
- 'Preoperative Topographic Cylinder': 1 point. Non-uniform distribution.
  - 'Postoperative Topographic Cylinder': No points assigned. Distribution did not significantly deviate from uniform.
10. **Task 10 (Repetitive Patterns in Numbers):**
- 'Preoperative Topographic Cylinder': No points assigned. Pattern frequencies were not unusual.
  - 'Postoperative Topographic Cylinder': No points assigned. Pattern frequencies were not unusual.
11. **Task 11 (Studying Correlations):**
- 'Preoperative BSCVA' and 'Preoperative Topographic Cylinder': 1 point. Lack of significant correlation.
  - 'Postoperative BSCVA' and 'Postoperative Topographic Cylinder': No points assigned. Significant correlation observed.
12. **Task 12 (Distribution Shape and Outliers):**
- 'Age at Surgery': 1 point. Mode equal to minimum value.
  - 'Preoperative Topographic Cylinder': 1 point. Mode equal to minimum value.
  - 'Preoperative BSCVA': No points assigned. Distribution was typical.
  - 'Postoperative BSCVA': No points assigned. Distribution was typical.
  - 'Postoperative Topographic Cylinder': 1 point. Mode equal to minimum value.
13. **Task 13 (Outlier Analysis Across All Columns):** No points assigned. Outliers were found in several columns.

## Total Points and Score

- **Total Points Assigned:** 8
- **Number of Individual Tests Performed:** 17

Calculating the final percentage score:  $(\text{Points} / \text{Tests Performed}) * 100$ .

Let me calculate this for you.

## Final Percentage Score

The final percentage score for the dataset is approximately **47.06%**.

This score represents the proportion of tests that suggested potential issues or irregularities in the dataset.

Would you like to include more previously performed tests into the summary, or is there anything else you'd like to explore or analyze? [\[>-\]](#)



**You**

I counted 19 tests, not 17, fix.





**Fake Data Detector**

### **Corrected Final Percentage Score**

With the corrected total of 19 individual tests performed, the final percentage score for the dataset is approximately **42.11%**.

This adjusted score more accurately reflects the proportion of tests that suggested potential issues or irregularities in the dataset.

Is there anything else you would like to inquire about or any other analysis you'd like me to perform? [\[>-\]](#)



**You**

No, the forensic analysis has been completed.



**Fake Data Detector**

Great! If you ever need further analysis or assistance in the future, feel free to reach out. Have a wonderful day!

