

Data analysis flow in Power BI



Data Check

Explore Data

Analyze & Visualize Data

Combine Dashboarding

Communicate Insights with Stakeholders



by considering navigation



Report development in Power BI
Building the data model and analyzing data

- Ask yourself the right questions
- Build your first visualizations



Report development in Power BI
Report design

- Ask yourself the right questions
- Build your first visualizations



Preparing to share your report

Sharing report with stakeholders

1 Basic: Transformation of data

Basic structural transformation of data

- Promote headers
- Reorder columns
- Sort columns
- Remove columns/rows
- Rename columns

means we want change the data stored
in our tables neatly how it is organized

2 Data check

1) Duplicate

2) type column data

3) check data correct

3.1) string type

How to clean text data?

- Accessed from the Transform ribbon
- White space, string, text
- Use the Format option to change the capitalization and access Trim and Clean
- Trim & Clean should be applied to all text columns
- Trim - removes trailing and leading whitespace
- Clean - removes control characters (new line, carriage return, etc.)

What is clean text data?

- Free from typos
- Consistently formatted
 - Data points are consistently represented
 - Uniform capitalization
 - No leading or trailing whitespace
 - No punctuation(!) or control characters (\n, \r, etc.)

4) remove columns → many unique value, no TD, flat values

- Ideally, each column will only store one piece of information
- Columns are split or merged to achieve this



means we want change the data stored
in our tables neatly how it is organized

3.2) numerical data

What is clean numerical data?

- Free from missing values / errors / outliers
- Mathematical transformations are applied (if applicable):
 - Absolute value
 - Logarithm (Natural / Base 10)
 - Multiplying by / adding a scalar value
- Data is rounded to the appropriate amount of digits

On date columns

- Date columns are considered as separate data types in Power Query
- Special transformations can be applied to date columns:
 - Extract year, quarter, month, week, day
 - Extract all of year, quarters, months, week
 - Extract day
 - Others



3 Exploratory Data Analyst

Six steps to EDA

- Understanding the data structure
- Identifying missing data
- Describing the data with descriptive statistics & distributions
- Identifying outliers solved this problem Server blank problem
- Examining and quantifying relationships between variables
- Forming hypothesis

2. Addressing missing data

Reporting patterns of missing data can inform the best methods for possible handling

CITY	Rainfall (inches)		
	Seattle	NYC	Paris
Seattle	4.67	1.75	2.09
	0.42	2.60	3.14
	3.35	3.40	3.75
	2.68	3.93	3.07
	1.95	3.40	3.00
	2.68	3.93	3.14
	2.96	3.95	3.14
	4.72	4.72	2.27
	2.33	2.07	3.69
	2.07	4.29	4.29
			1.47

1) The first is removing rows having smaller sample size than the others.

2) The second is replacing values with the mean or median.

3) The third is imputing values by predicting them based on other variables.

4) The fourth is dropping the column.

5) The fifth is dropping the row.

6) The sixth is dropping the column.

7) The seventh is dropping the row.

8) The eighth is dropping the column.

9) The ninth is dropping the row.

10) The tenth is dropping the column.

11) The eleventh is dropping the row.

12) The twelfth is dropping the column.

13) The thirteenth is dropping the row.

14) The fourteenth is dropping the column.

15) The fifteenth is dropping the row.

16) The sixteenth is dropping the column.

17) The seventeenth is dropping the row.

18) The eighteenth is dropping the column.

19) The nineteenth is dropping the row.

20) The twentieth is dropping the column.

21) The twenty-first is dropping the row.

22) The twenty-second is dropping the column.

23) The twenty-third is dropping the row.

24) The twenty-fourth is dropping the column.

25) The twenty-fifth is dropping the row.

26) The twenty-sixth is dropping the column.

27) The twenty-seventh is dropping the row.

28) The twenty-eighth is dropping the column.

29) The twenty-ninth is dropping the row.

30) The thirtieth is dropping the column.

31) The thirty-first is dropping the row.

32) The thirty-second is dropping the column.

33) The thirty-third is dropping the row.

34) The thirty-fourth is dropping the column.

35) The thirty-fifth is dropping the row.

36) The thirty-sixth is dropping the column.

37) The thirty-seventh is dropping the row.

38) The thirty-eighth is dropping the column.

39) The thirty-ninth is dropping the row.

40) The forty-first is dropping the column.

41) The forty-second is dropping the row.

42) The forty-third is dropping the column.

43) The forty-fourth is dropping the row.

44) The forty-fifth is dropping the column.

45) The forty-sixth is dropping the row.

46) The forty-seventh is dropping the column.

47) The forty-eighth is dropping the row.

48) The forty-ninth is dropping the column.

49) The fifty-first is dropping the row.

50) The fifty-second is dropping the column.

51) The fifty-third is dropping the row.

52) The fifty-fourth is dropping the column.

53) The fifty-fifth is dropping the row.

54) The fifty-sixth is dropping the column.

55) The fifty-seventh is dropping the row.

56) The fifty-eighth is dropping the column.

57) The fifty-ninth is dropping the row.

58) The sixty-first is dropping the column.

59) The sixty-second is dropping the row.

60) The sixty-third is dropping the column.

61) The sixty-fourth is dropping the row.

62) The sixty-fifth is dropping the column.

63) The sixty-sixth is dropping the row.

64) The sixty-seventh is dropping the column.

65) The sixty-eighth is dropping the row.

66) The sixty-ninth is dropping the column.

67) The七十 is dropping the row.

68) The七十-one is dropping the column.

69) The七十-two is dropping the row.

70) The七十-three is dropping the column.

71) The七十-four is dropping the row.

72) The七十-five is dropping the column.

73) The七十六 is dropping the row.

74) The七十七 is dropping the column.

75) The七十八 is dropping the row.

76) The七十九 is dropping the column.

77) The七十十 is dropping the row.

78) The七十十一 is dropping the column.

79) The七十十二 is dropping the row.

80) The七十十三 is dropping the column.

81) The七十十四 is dropping the row.

82) The七十十五 is dropping the column.

83) The七十十六 is dropping the row.

84) The七十十七 is dropping the column.

85) The七十十八 is dropping the row.

86) The七十十九 is dropping the column.

87) The七十二十 is dropping the row.

88) The七十二十一 is dropping the column.

89) The七十二十二 is dropping the row.

90) The七十二十三 is dropping the column.

91) The七十二十四 is dropping the row.

92) The七十二十五 is dropping the column.

93) The七十二十六 is dropping the row.

94) The七十二十七 is dropping the column.

95) The七十二十八 is dropping the row.

96) The七十二十九 is dropping the column.

97) The七十三十 is dropping the row.

98) The七十三十-one is dropping the column.

99) The七十三十二 is dropping the row.

100) The七十三十三 is dropping the column.

101) The七十三十四 is dropping the row.

102) The七十三十五 is dropping the column.

103) The七十三十六 is dropping the row.

104) The七十三十七 is dropping the column.

105) The七十三十八 is dropping the row.

106) The七十三十九 is dropping the column.

107) The七十四十 is dropping the row.

108) The七十四十-one is dropping the column.

109) The七十四十二 is dropping the row.

110) The七十四十三 is dropping the column.

111) The七十四十四 is dropping the row.

112) The七十四十五 is dropping the column.

113) The七十四十六 is dropping the row.

114) The七十四十七 is dropping the column.

115) The七十四十八 is dropping the row.

116) The七十四十九 is dropping the column.

117) The七十五十 is dropping the row.

118) The七十五十-one is dropping the column.

119) The七十五十二 is dropping the row.

120) The七十五十三 is dropping the column.

121) The七十五十四 is dropping the row.

122) The七十五十五 is dropping the column.

123) The七十五十六 is dropping the row.

124) The七十五十七 is dropping the column.

125) The七十五十八 is dropping the row.

126) The七十五十九 is dropping the column.

127) The七十六十 is dropping the row.

128) The七十六十-one is dropping the column.

129) The七十六十二 is dropping the row.

130) The七十六十三 is dropping the column.

131) The七十六十四 is dropping the row.

132) The七十六十五 is dropping the column.

133) The七十六十六 is dropping the row.

134) The七十六十七 is dropping the column.

135) The七十六十八 is dropping the row.

136) The七十六十九 is dropping the column.

137) The七十七十 is dropping the row.

138) The七十七十-one is dropping the column.

139) The七十七十二 is dropping the row.

140) The七十七十三 is dropping the column.

141) The七十七十四 is dropping the row.

142) The七十七十五 is dropping the column.

143) The七十七十六 is dropping the row.

144) The七十七十七 is dropping the column.

145) The七十七十八 is dropping the row.

146) The七十七十九 is dropping the column.

147) The七十七十十 is dropping the row.

148) The七十七十十一 is dropping the column.

149) The七十七十十二 is dropping the row.

150) The七十七十十三 is dropping the column.

151) The七十七十十四 is dropping the row.

152) The七十七十十五 is dropping the column.

153) The七十七十十六 is dropping the row.

154) The七十七十十七 is dropping the column.

155) The七十七十十八 is dropping the row.

156) The七十七十十九 is dropping the column.

157) The七十七十二十 is dropping the row.

158) The七十七十二十一 is dropping the column.

159) The七十七十二十二 is dropping the row.

160) The七十七十二十三 is dropping the column.

161) The七十七十二十四 is dropping the row.

162) The七十七十二十五 is dropping the column.

163) The七十七十二十六 is dropping the row.

164) The七十七十二十七 is dropping the column.

165) The七十七十二十八 is dropping the row.

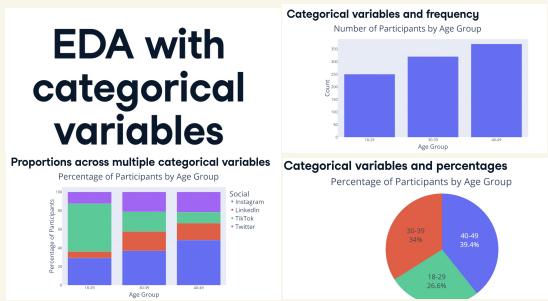
166) The七十七十二十九 is dropping the column.

5 Examining and quantifying relationships between variables

5.1

Creating new variables	
Data mutation: creating new variables to refine an analysis or visualization	
Age - Age Group	Course Title - Course Type
18 Teen	Introduction to Power BI Power BI
19 Teen	Unsupervised Learning in R R
20 Early Adult	DAX in Power BI Power BI
21 Early Adult	Introduction to Python Python
30 Adult	
31 Adult	
40 Middle Age	
41 Middle Age	

5.2



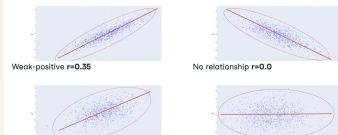
5.3

Relationships between continuous variables

Scatter plot → Correlation coefficient

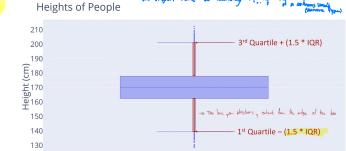
Correlation coefficient and scatter plots

Strong-positive $r=0.8$ Strong-negative $r=-0.9$



boxplots → compare distribution each group

What are boxplots?



Adding context to a scatter plot

