Complete Beginner Foundation Assessment

250+ Questions to Build Rock-Solid Data Science & Analytics Foundation

---

PART 1: BASIC STATISTICS FOUNDATION (40 Questions)

Measures of Central Tendency

1. What is the mean and how do you calculate it?
2. What is the median and when is it better than mean?
3. What is the mode and when do you use it?
4. If you have values [2, 4, 6, 8, 100], which measure (mean/median) is better and why?
5. Can a dataset have more than one mode?
6. How do outliers affect mean vs median?
7. What is the weighted average?
8. When would you use geometric mean instead of arithmetic mean?
9. What is the midrange of a dataset?
10. How do you interpret the median in real-world scenarios?

Measures of Spread

11. What is range and how do you calculate it?
12. What is variance?
13. What is standard deviation?
14. Why is standard deviation more useful than variance?
15. What does a small standard deviation tell you?
16. What does a large standard deviation indicate?
17. What is the interquartile range (IQR)?
18. What are quartiles (Q1, Q2, Q3)?
19. How do you identify outliers using IQR?
20. What is coefficient of variation?

Distributions & Probability

21. What is a normal distribution (bell curve)?

22. What does it mean when data is "normally distributed"?
23. What is the 68-95-99.7 rule?
24. What is a skewed distribution?
25. What is positive skew vs negative skew?
26. What is a uniform distribution?
27. What is probability (in simple terms)?
28. If you flip a fair coin, what's the probability of heads?
29. What does "independent events" mean?
30. What is the difference between "and" vs "or" in probability?

Basic Statistical Concepts

31. What is a population in statistics?
32. What is a sample?
33. Why do we use samples instead of studying entire populations?
34. What is sampling bias?
35. What is random sampling?
36. What is a percentile?
37. If you score in the 90th percentile, what does that mean?
38. What is correlation?
39. Can correlation be negative? What does that mean?
40. What does "correlation does not imply causation" mean with an example?

PART 2: PYTHON FUNDAMENTALS (50 Questions)

Basic Data Types & Variables

41. What is a variable in Python?
42. How do you assign a value to a variable?
43. What is an integer (int)?
44. What is a float?
45. What is a string?
46. What is a boolean?
47. How do you check the type of a variable?
48. What happens when you add two strings together?
49. What is the difference between '5' and 5?
50. How do you convert a string to an integer?

Lists

51. What is a list in Python?
52. How do you create an empty list?

53. How do you access the first element of a list?
54. What is indexing in lists?
55. What does index -1 mean?
56. How do you add an element to the end of a list?
57. How do you remove an element from a list?
58. What is slicing in lists?
59. How do you find the length of a list?
60. What does `list[2:5]` return?
61. How do you check if an item exists in a list?
62. How do you sort a list?
63. What is the difference between sort() and sorted()?
64. How do you reverse a list?
65. How do you concatenate two lists?

Dictionaries

66. What is a dictionary in Python?
67. How is a dictionary different from a list?
68. How do you create a dictionary?
69. How do you access a value in a dictionary?
70. How do you add a new key-value pair?
71. How do you check if a key exists in a dictionary?
72. How do you get all keys from a dictionary?
73. How do you get all values from a dictionary?
74. What happens if you try to access a key that doesn't exist?
75. How do you remove a key-value pair from a dictionary?

Control Flow

76. What is an if statement?
77. What is an else statement?
78. What is elif?
79. How do comparison operators work (==, !=, <, >, <=, >=)?
80. What are logical operators (and, or, not)?
81. What is a for loop?
82. How do you loop through a list?
83. What is the range() function?
84. What is a while loop?
85. When would you use a while loop instead of a for loop?
86. What is the break statement?
87. What is the continue statement?
88. What is nested loop?
89. How do you loop through a dictionary?
90. What is list comprehension (basic)?

PART 3: PANDAS BASICS (50 Questions)

DataFrames & Series Basics

91. What is Pandas library used for?
92. How do you import Pandas?
93. What is a DataFrame?
94. What is a Series?
95. How is a DataFrame different from a dictionary?
96. How do you create a DataFrame from a dictionary?
97. How do you create a DataFrame from a list of lists?
98. What is the difference between a DataFrame and a spreadsheet?
99. How do you display the first 10 rows of a DataFrame?
100. How do you display the last 5 rows?

Reading & Exploring Data

101. How do you read a CSV file into a DataFrame?
102. How do you read an Excel file?
103. What does df.shape tell you?
104. What does df.columns return?
105. How do you get the list of column names?
106. What does df.info() show you?
107. What does df.describe() do?
108. What statistics does describe() provide?
109. How do you check the data type of each column?
110. How do you view the index of a DataFrame?

Selecting Data

111. How do you select a single column?
112. How do you select multiple columns?
113. What is the difference between df['column'] and df[['column']]?
114. What does loc do?
115. What does iloc do?
116. What is the difference between loc and iloc?
117. How do you select the first row using iloc?
118. How do you select rows 5 to 10?
119. How do you select specific rows and columns together?
120. How do you select rows based on a condition?

Data Cleaning Basics

121. How do you check for missing values?
122. What does isnull() return?
123. How do you count missing values in each column?
124. How do you drop rows with missing values?
125. How do you drop columns with missing values?
126. How do you fill missing values with a specific value?
127. How do you fill missing values with the mean?
128. How do you drop duplicate rows?
129. How do you identify duplicate rows?
130. How do you rename a column?

Basic Data Operations

131. How do you sort a DataFrame by a column?
132. How do you sort in descending order?
133. How do you filter rows where a column value > 50?
134. How do you filter rows with multiple conditions?
135. What does groupby() do?
136. How do you calculate the mean of groups?
137. How do you count the number of rows in each group?
138. How do you reset the index?
139. How do you set a column as the index?
140. How do you add a new column to a DataFrame?

PART 4: DATA VISUALIZATION BASICS (40 Questions)

Why & When to Visualize

141. Why is data visualization important?
142. What are the benefits of visualizing data over tables?
143. When should you use a chart vs showing raw numbers?
144. What is exploratory data analysis (EDA)?
145. How do visualizations help in presentations?
146. What makes a good data visualization?
147. What makes a bad data visualization?
148. What is chart junk?
149. Why should every chart have a title?
150. Why are axis labels important?

Chart Types & Usage

151. What is a bar chart used for?
152. What is a histogram?
153. What is the difference between a bar chart and histogram?
154. When would you use a line chart?
155. What is a scatter plot?
156. What does a scatter plot help you see?
157. What is a pie chart best used for?
158. Why are pie charts sometimes criticized?
159. What is a box plot (box and whisker plot)?
160. What information does a box plot show?
161. What is a heatmap?
162. When would you use a heatmap?
163. What is a stacked bar chart?
164. What is a horizontal bar chart?
165. When would you use horizontal bars instead of vertical?

Matplotlib Basics

166. What is Matplotlib?
167. How do you import Matplotlib?
168. What is plt.plot() used for?
169. How do you create a simple line plot?
170. How do you add a title to a plot?
171. How do you label the x-axis?
172. How do you label the y-axis?
173. How do you change the color of a plot?
174. How do you change the line style?
175. How do you add a legend?
176. How do you create a bar chart?
177. How do you create a scatter plot?
178. How do you create a histogram?
179. How do you save a plot as an image?
180. How do you show multiple plots in one figure?

Seaborn Basics

181. What is Seaborn?
182. How is Seaborn different from Matplotlib?
183. How do you import Seaborn?
184. How do you create a scatter plot with Seaborn?
185. What is sns.countplot() used for?
186. What is sns.boxplot() used for?
187. How do you create a heatmap with Seaborn?
188. What is sns.pairplot()?

189. How do you change the color palette in Seaborn?
190. Why use Seaborn over Matplotlib?

PART 5: SQL FUNDAMENTALS (50 Questions)

Database Basics

191. What is SQL?
192. What is a database?
193. What is a table in a database?
194. What is a row in a table?
195. What is a column in a table?
196. What is a record?
197. What is a field?
198. What are the basic SQL commands?
199. What is a query?
200. What is the difference between SQL and MySQL/PostgreSQL?

SELECT Statement

201. What does the SELECT statement do?
202. How do you select all columns from a table?
203. How do you select specific columns?
204. How do you select unique values?
205. What does DISTINCT do?
206. How do you limit the number of rows returned?
207. What does the LIMIT clause do?
208. How do you select the top 5 rows?
209. Can you select columns in any order?
210. How do you give a column an alias (nickname)?

Filtering Data

211. What is the WHERE clause?
212. How do you filter rows where age > 25?
213. How do you filter rows where name = 'John'?
214. What is the difference between = and ==?
215. How do you use multiple conditions with AND?
216. How do you use multiple conditions with OR?
217. What does the BETWEEN operator do?
218. How do you check if a value is in a list (IN operator)?
219. What does the LIKE operator do?

220. How do you search for patterns using LIKE and wildcards?
221. What does % wildcard mean?
222. What does _ wildcard mean?
223. How do you filter for NULL values?
224. How do you filter for NOT NULL values?
225. What is the NOT operator?

Sorting & Aggregating

226. What does ORDER BY do?
227. How do you sort in ascending order?
228. How do you sort in descending order?
229. How do you sort by multiple columns?
230. What is COUNT()?
231. What is SUM()?
232. What is AVG()?
233. What is MIN()?
234. What is MAX()?
235. What does GROUP BY do?
236. How do you count rows in each group?
237. How do you calculate average by group?
238. What is the HAVING clause?
239. What is the difference between WHERE and HAVING?
240. How do you filter groups after aggregation?

PART 6: FOUNDATIONAL CONCEPTS (30 Questions)

Data Types & Formats

241. What is structured data?
242. What is unstructured data?
243. Give examples of structured data
244. Give examples of unstructured data
245. What is a CSV file?
246. What is JSON format?
247. What is an Excel file (.xlsx)?
248. What is a text file (.txt)?
249. What is tabular data?
250. What is semi-structured data?

Data Analysis Process

251. What are the steps in a typical data analysis project?
252. What is data collection?
253. What is data cleaning?
254. Why is data cleaning important?
255. What is data exploration?
256. What is data analysis?
257. What is data visualization in the process?
258. What is communicating results?
259. What is a data pipeline?
260. What is the difference between data and information?

Key Terminology

261. What is a dataset?
262. What is a variable in data analysis?
263. What is an observation?
264. What is a feature?
265. What is a target variable?
266. What is data quality?
267. What is data integrity?
268. What is data consistency?
269. What is data accuracy?
270. What is data completeness?

SCORING GUIDE

Points per Question:

- **Fully understand & can explain**: 3 points
- **Partially understand**: 2 points
- **Heard of it but unsure**: 1 point
- **Don't know**: 0 points

Total Score Interpretation:

- **0-200**: Just starting - Focus on basics first
- **201-400**: Building foundation - Keep practicing
- **401-550**: Good progress - Start small projects
- **551-650**: Solid foundation - Ready for intermediate
- **651-750**: Strong foundation - Ready for real projects
- **751-810**: Excellent foundation - Move to advanced topics

---

RECOMMENDED LEARNING PATH BY SECTION

If Score Low in Statistics (Q1-40):

📚 **Resources:**

- Khan Academy Statistics
- StatQuest YouTube channel
- "Naked Statistics" book
- Practice with real datasets

🎯 **Practice:**

- Calculate mean/median/mode manually
- Identify distributions in real data
- Create examples of correlation

If Score Low in Python (Q41-90):

📚 **Resources:**

- Python.org tutorial
- Codecademy Python course
- "Automate the Boring Stuff" book
- LeetCode easy problems

🎯 **Practice:**

- Write 10-15 lines of code daily
- Solve simple problems
- Create small programs
- Practice data type conversions

If Score Low in Pandas (Q91-140):

📚 **Resources:**

- Pandas official documentation
- Kaggle Pandas course

- "Python for Data Analysis" book
- Practice datasets on Kaggle

🎯 **Practice:**

- Load and explore 5 different datasets
- Perform cleaning operations
- Practice filtering and grouping
- Create summary statistics

If Score Low in Visualization (Q141-190):

📚 **Resources:**

- Matplotlib official tutorials
- Seaborn documentation
- "Storytelling with Data" book
- DataViz projects on Kaggle

🎯 **Practice:**

- Create 1 chart daily
- Recreate charts you see online
- Visualize your own data
- Try different chart types

If Score Low in SQL (Q191-240):

📚 **Resources:**

- SQLZoo
- Mode Analytics SQL tutorial
- LeetCode SQL problems
- W3Schools SQL

🎯 **Practice:**

- Write 5 queries daily
- Use sample databases
- Practice filtering and grouping
- Solve SQL challenges

If Score Low in Concepts (Q241-270):

📚 **Resources:**

- Introduction to Data Science courses
- Data analysis blogs
- YouTube data science channels
- LinkedIn Learning

🎯 **Practice:**

- Read data analysis case studies
- Follow data professionals
- Join data communities
- Practice explaining concepts

30-DAY FOUNDATION BUILDING PLAN

Week 1: Statistics & Python Basics

- Day 1-3: Central tendency & spread
- Day 4-5: Python data types & variables
- Day 6-7: Lists and dictionaries

Week 2: Python Control Flow & Pandas Start

- Day 8-10: If statements and loops
- Day 11-12: Functions basics
- Day 13-14: Introduction to Pandas

Week 3: Pandas Operations & Visualization

- Day 15-17: Data selection and filtering
- Day 18-19: Data cleaning basics
- Day 20-21: Basic plots with Matplotlib

Week 4: SQL & Integration

- Day 22-24: SELECT and WHERE
- Day 25-26: GROUP BY and aggregations

- Day 27-28: Practice mini-project
- Day 29-30: Review and portfolio piece

PRACTICE PROJECTS FOR BEGINNERS

Project 1: Sales Data Analysis

- Load sales CSV
- Clean missing values
- Calculate total sales by category
- Create bar chart
- Find top products

Project 2: Student Grade Analysis

- Import student data
- Calculate class average
- Find median score
- Identify outliers
- Visualize grade distribution

Project 3: Weather Data Exploration

- Load temperature data
- Calculate monthly averages
- Find hottest/coldest days
- Create line chart
- Identify trends

Project 4: Customer Database Query

- Create simple database
- Write SELECT queries
- Filter by conditions
- Group customers by city
- Count customers per region

Project 5: Survey Data Visualization

- Clean survey responses
- Count responses per category
- Calculate percentages

- Create pie chart
- Make bar chart

COMMON BEGINNER MISTAKES TO AVOID

❌ **Jumping to ML without fundamentals** ✅ Master basics first

❌ **Not practicing coding daily** ✅ Code for at least 30 minutes daily

❌ **Memorizing without understanding** ✅ Understand concepts deeply

❌ **Working without real data** ✅ Use real datasets from day one

❌ **Learning tools without statistics** ✅ Balance theory and tools

❌ **Not documenting code** ✅ Add comments and explanations

❌ **Comparing yourself to others** ✅ Track your own progress

❌ **Giving up when stuck** ✅ Debugging is part of learning

NEXT STEPS AFTER COMPLETING THIS

✅ **If scored 650+**: Move to intermediate questions ✅ **If scored 400-649**: Complete 3-5 beginner projects ✅ **If scored below 400**: Focus on weak areas first

Remember: Everyone starts somewhere. The key is consistent daily practice! 🚀