

SKILLS CHECKLIST

DATA ANALYST IN 2025

VARUN SAGAR THEEGALA

THE LIST

As a Data Analyst, you'll be expected to possess the following toolkit at minimum –

- **STATISTICS**
- **MICROSOFT EXCEL**
- **SQL**
- **PYTHON**
- **POWER BI**
- **DATA STORYTELLING**

Let's dive deeper into each skill and see what you need to prepare ➡

STATISTICS

Helps interpret data, derive insights, and make data-driven decisions

- **Descriptive Statistics:** Mean, median, mode, variance, standard deviation.
- **Distributions:** Normal, binomial, and uniform for understanding data patterns.
- **Correlation vs. Causation:** Correlation coefficients and identifying relationships.
- **Outlier Detection:** Z-scores, IQR, and visual methods like box plots.
- **Hypothesis Testing:** t-tests, chi-square tests, p-values for decision-making.
- **Confidence Intervals:** Estimating ranges for metrics with certainty.

MS EXCEL

Essential for data manipulation, analysis & visualization in spreadsheets.

- **Data Cleaning:** Removing duplicates, handling missing data.
- **Pivot Tables:** Summarizing and analyzing datasets.
- **Formulas & Functions:** VLOOKUP, INDEX-MATCH, SUMIF, COUNTIF, etc.
- **Charts & Visualizations:** Bar, pie, line charts; conditional formatting.
- **Data Validation:** Restricting input with dropdowns and rules.
- **Power Query:** Automating data cleaning and transformation tasks.

STRUCTURED QUERY LANGUAGE (SQL)

Crucial for querying and managing large datasets in databases.

- **Data Retrieval:** SELECT, WHERE, ORDER BY.
- **Joins:** INNER, LEFT, RIGHT, and FULL JOINS.
- **Aggregations:** GROUP BY, HAVING, COUNT, SUM, AVG.
- **Subqueries & CTEs:** Writing nested queries and temporary tables.
- **Window Functions:** RANK, DENSE_RANK, ROW_NUMBER, LAG, LEAD.
- **Database Design:** Normalization and schema creation.
- **Optimization:** Indexing, query performance tuning.

PYTHON

Automates tasks, processes data, and builds analytical models efficiently.

- **Data Manipulation:** Pandas for DataFrames and series operations.
- **Data Visualization:** Matplotlib, Seaborn, Plotly for interactive graphs.
- **Numerical Computing:** NumPy for mathematical operations.
- **APIs & Web Scraping:** Using requests, BeautifulSoup, or Scrapy.
- **Automation:** Automating tasks with Python scripts.
- **Data Cleaning:** Handling missing data, regex for text data.
- **Statistics:** Scipy

POWER BI

Enables interactive dashboards and reports for business decision-making.

- **Data Modeling:** Relationships, calculated columns, measures.
- **DAX:** Functions like CALCULATE, SUMX, and RANKX.
- **Visualizations:** Bar charts, maps, scatter plots, KPI indicators.
- **Advanced Features:** Drill-through, hierarchies, bookmarks.
- **Power Query:** Data transformation and loading.
- **Publishing Reports:** Using Power BI Service and workspaces.
- **Row-Level Security (RLS):** Securing sensitive data.

DATA STORYTELLING

Communicate complex data with an engaging narratives for clear insights.

- **Structuring Insights:** Understanding your audience and their needs.
- **Chart Selection:** Picking the right chart for your message.
- **Highlighting Trends:** Using annotations and visual cues.
- **Narrative Flow:** Beginning with context, ending with actionable insights.
- **Avoiding Overload:** Keeping dashboards simple and focused.
- **Contextual Formatting:** Using colors and labels wisely.
- **Communication Skills:** Presenting insights confidently and concisely.

EARN BROWNIE POINTS

**ADDITIONAL
SKILLS
GOOD TO HAVE**

AI TOOLS (CHATGPT)*

Enhance efficiency by automating tasks, troubleshooting & aid learning

- **Prompt Engineering:** Craft precise prompts for accurate outputs.
- **Iterative Questioning:** Break queries into smaller questions to refine answers.
- **Code Generation:** Generate code snippets for SQL & Python tasks.
- **Code Debugging:** Troubleshoot SQL, Python, and other code issues.
- **Documentation:** Create clear analysis reports or code comments.
- **Learning Aid:** Clarify doubts & accelerate tool learning.
- **Professional Communication:** Refine presentations and emails for impact.

CLOUD COMPUTING

Scalable storage and computing power for handling large datasets.

- **Cloud Storage:** Store and access data on cloud platforms.
- **Cloud Databases:** Query cloud-hosted databases.
- **Data Pipelines:** Automate data processing tasks.
- **ETL on Cloud:** Extract, transform, and load data to the cloud.
- **Cloud Security:** Manage data privacy and access control.
- **Cost Management:** Monitor cloud usage and costs.

GITHUB

Enables version control and collaboration on data analysis projects.

- **Version Control:** Understand Git basics—repositories, commits, and the staging area—to manage project changes.
- **Branching and Merging:** Learn to create branches for new features and merge them back into the main codebase.
- **Pull Requests:** Proposing, reviewing, and discussing changes before integration.
- **Conflict Resolution:** Develop skills to identify and resolve merge conflicts when multiple contributors edit the same content.
- **Collaborative Workflows:** Adopt best practices like clear commit messages and code reviews to maintain project quality.

CONCLUSION

Embrace continuous learning and growth to remain competitive in the ever-changing data landscape.

Ensure you **possess a strong hold over the basic toolkit** & and you're **ready to apply & appear for interviews**.

Go the extra steps to **learn trending skills** like AI and Cloud and **own a competitive edge**.