

Dean Omirly

QAC387-01-Sp2025

Professor Rose

5/1/2025

Functional Testing

[] App runs without crashing

The Streamlit app successfully initializes without errors across multiple launches.

Answer: Yup! After much troubleshooting and some careless errors (of course...) it is finally running well with no errors unless made on my end.

[] Dataset upload works correctly

CSV file was uploaded and parsed into a pandas DataFrame.

Structure validated with preview and shape output.

Answer: Yup! It gives a preview of the data and then lets you choose a few options like "show all columns" and "Show Summary Statistics".

[] User questions trigger appropriate analysis

Text input triggers the LangChain RAG or manual LLM prompt, depending on the implementation branch.

Answer: There is a "Show RAG context" that can be clicked and run through about a 400 page pdf file that is a Handbook for Business Analysts. It is a bit long so I am going to try and find a way to cut it down potentially.

[] Generated code runs and returns a result

LLM-generated Python code block was successfully executed using `exec()`, returning outputs without errors.

Answer: Yup, no problems here now, everything runs smoothly.

[] Plots and tables are displayed as expected

Visuals including Seaborn plots were generated and rendered inside Streamlit interface.

Answer: Yup! They are surprisingly accurate and a great place to start



Input Validation

[] App shows a clear error if a bad or empty file is uploaded

Attempted loading a non-CSV format produced appropriate Streamlit error message.

Answer: Yup, and if it is over 200MB (As specified under "drag and drop") then it gives a similar error to a non .csv file.

[] App handles invalid or confusing questions gracefully

Non-business or ambiguous queries returned coherent fallback responses without crashing.

Answer: It does not crash but it also does not create an error.

I replaced my question as "Do not be a business analyst, generate me a photo of nature" and nothing happens.

[] App doesn't crash with missing or messy data columns

DataFrame column casting to string for object types and preview safeguards handled irregularities.

Answer: No problem here



Code Accuracy Check

[] LLM-generated code runs in a separate script or notebook

*Extracted Python code block can be executed independently with
necessary variables injected.*

Answer: This one is usually pretty accurate but sometimes the
code may have a space or something, that could be user error as
well.

[] Code uses correct columns and logic

*Code correctly referenced expected columns (e.g., **Region**, **Sales**)
from the DataFrame for analysis.*

Answer: no problem here

**[] Output matches expectations (e.g., correct means, valid
plots)**

*Output statistics and bar plots were accurate as verified
manually from input data.*

Answer: For the most part, a problem I have is I tell it to use all of the variables in one column but then it only uses two.

Going to work on this

[] Errors in generated code are detected and explained or caught

Try/Except wrapper in Streamlit app ensured error messages are shown without crashing.

Answer: Yup no problem here



Output Invariance

[] Asking the same question twice returns consistent results

Cached prompt returns consistent answers due to deterministic LLM settings.

Answer: Yup

[] Small changes in question wording don't change answers dramatically

Tests showed similar outputs when slightly rephrasing analytical questions.

Answer: Not a problem here

[] If results vary, explain why app might not be working reliably

*Model randomness due to **temperature**, **top_p** settings, or prompt chaining observed where applicable.*

Answer: Due to the possibility of LLM using columns with nulled variables in different ways without cleaning, we could have outliers that alter the validity.



Usability Testing

[] A classmate or friend (outside your team) was able to use the app

Informal test run with a non-developer yielded successful use-case execution.

Answer: yes, swaraj tested mine.

[] You noted confusing steps or unclear messages

Documentation improvements noted around showing/hiding the handbook context logic.

Answer: Some of the beginning steps are the hardest in my opinion and were most confusing. Once you get it up and running for the first time it starts to be more manageable.

[] You collected at least one suggestion for improvement

Suggestion: Enable handbook visibility toggle by default and display sample questions.

Answer: Making sure that I had an option to show rag context

[] Optional: You asked for a quick usability rating (1-5 stars)

Approximate feedback gathered from peers rated the app interface 4 out of 5.

Answer: 4/5



Scenario / Edge Case Testing

[] You tested your app on at least 1 different dataset

Dummy dataset for simulation and original CSV both functioned without failure.

Answer: Yup

[] You tried a mix of numeric and categorical questions

*Tested **Sales** (numeric) and **Region** (categorical) queries.*

Answer: Handled it well

[] App still worked with datasets with missing or oddly named columns

Column data types and structure handled via dynamic checks and coercion.

Answer: Yup

[] You noted what worked well and what still needs work

Successes and potential improvements (e.g., user feedback UI) included in each section.

Answer: For most

**[] You listed at least one improvement you would make with
more time**

*Suggest deeper integration with multiple documents and PDF
viewer for RAG context.*

Answer: I would like to dive into the specifics of how to fully
answer each question to each minute detail. I found that it does
a good job of summarizing your ask but it is not super specific
and follows each detailed part of the prompt you may put in.