**Google Data Analytics Professional Certificate – Coursera**

Intro: This is a documentation of my process as I completed the capstone project included in the online courser course: Google Data Analytics Professional Certificate. This project situation had me working as a junior data analyst for a business intelligence consultant. I have been at my job for six months, and my boss feels I am ready for more responsibility. He has asked me to lead a project for a brand new client — this will involve everything from defining the business task all the way through presenting my data-driven recommendations. I will choose the topic, ask the right questions, identify a fresh dataset and ensure its integrity, conduct analysis, create compelling data visualizations, and prepare a presentation.

**Ask**: Five questions that must be answered by my team during this process are

1. What type of company does my client represent, and what are they asking me to accomplish?

The company I represent is heavily involved in the gasoline industry and they wanted me to compare the recent gas averages per gallon across the country to previous years.

2. What are the key factors involved in the business task I am investigating?

The types of gas, currency, time, state averages, and conflicting current events.

3. What type of data will be appropriate for my analysis?

The average price of gasoline from the last 30 years, states average price, and gas types average price

4. Where will I obtain that data?

I used a combination of open souce .csv files and webscrapping from the following sites.

<https://fred.stlouisfed.org/series/GASREGW>

<https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=EMM_EPMR_PTE_NUS_DPG&f=W>

<https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=pet&s=emm_epmr_pte_nus_dpg&f=m>

<https://gasprices.aaa.com/state-gas-price-averages/>

This was all conducted (03/04/2022)

5. Who is my audience, and what materials will help I present to them effectively?

The audience I will be presenting to is a client and company involved in gasoline products so they want to find out how the recent trends will affect overall production

Question 1. Subtasks

I will produce a report with the following deliverables:

1. A clear statement of the business task I have selected to investigate

I was tasked with tracking the price per gallon of gasoline using the average price per gallon, states regular gas price average, and states average gas price among four types of gas. From this I needed to identify a trend and make a business plan

2. A description of all data sources used

The data I am using is a collection of open source data from U.S gas sales over the course of 30 years

3. Documentation of any cleaning or manipulation of data

Most of the data was already cleaned fairly well, but I had to reconstruct some of the missing values from the dataset.

4. A summary of my analysis

The average price of gas is on a upward trend that is more than likely going to break all-time records, this type of increase is going to have an impact on all scales.

5. Supporting visualizations and key findings

I created a tableau data visualization to show how the price per gallon is almost at an all-time high.

6. Based on what my discover, a list of additional deliverables I think would be helpful to include for further exploration

Current events that are driving the current price increase. Russia gas ban and overall Ukraine conflict is root cause.

7. My top high-level insights based on the analysis

Gasoline is about to reach an all-time high and the costs of anything involving gasoline will increase in expense so it would make sense to re-analyze specific aspects of a company that will be impacted.

Guiding questions:

What topic am I exploring?

The price per gallon of gas increase in comparison to the last 30 years.

What is the problem I am trying to solve?

Why and what is the trend for gas prices seeing an increase for?

What metrics will I use to measure my data to achieve my objective?

U.S.D, Price Per Gallon, Average over 30 years.

Who are the stakeholders?

Employees, Gasoline Industry Representatives, U.S Gas Consumers.

Who is my audience?

The client and boss that I am presenting my findings to for the average price per gallon of gas increase.

How can I insights help my client make decisions?

They may decide to configure their production or purchasing amount with the current price of U.S.D, in a real-world situation product cost and the expense of gas would have factors for many of the company choices.

Key tasks:

1. Identify the business task

Investigate the price per gallon of gas and how this increase will impact company performance and expense.

2. Consider key stakeholders

Are these stakeholders benefitting from higher gas prices by selling or are they being hurt by this price increase?

\*Deliverable\*

A clear statement of the business task I have selected to investigate

**Prepare**:

Guiding questions:

Where is my data located?

The data is saved locally, and I have a GitHub / Tableau public holding the information for those who want to see the details.

How is the data organized?

This data is organized by average of each state, the u.s average gas price, and the gasregw over 30 years.

Are there issues with bias or credibility in this data?

No, this is open source data from the U.S average gas and is creditable.

Does my data ROCCC?

Yes, this data meets Reliable, Original, Comprehensive, Current, and Cited.

How am I addressing licensing, privacy, security, and accessibility?

This is all open source and has no privacy issues as it has no impact on current events, but rather a summary of the past.

How did I verify the data’s integrity?

I compared the data I used to other forms to make sure that the dates were matching with the price average and the data was correct.

How does it help me answer my question?

Having reliable and accurate data helps make forcasting.

Are there any problems with the data?

Yes, the month of december in 1990 is missing values for six separate days.

Key tasks:

1. Download data and store it appropriately.

2. Identify how it’s organized.

3. Sort and filter the data.

4. Determine the credibility of the data.

\*Deliverable\*

A description of all data sources used

**Process:**

Guiding questions:

What tools am I choosing and why?

I used excel for cleaning the data and to gain a better visual understanding of it, I then took it into tableau to make it look appealing.

Have I ensured my data’s integrity?

Yes

What steps have I taken to ensure that my data is clean?

Using open source data that is comparative to similar data as well as other forms of data.

How can I verify that my data is clean and ready to analyze?

I made sure each column had an appropriate value for the assigned date, those that did not I tried searching for the correct information but decided to omit it entirely since I would prefer to have accurate data.

Have I documented the cleaning process so I can review and share those results?

Yes

Key tasks:

1. Check the data for errors.

2. Choose my tools.

3. Transform the data so I can work with it effectively.

4. Document the cleaning process.

\*Deliverable\*

Documentation of any cleaning or manipulation of data

**Analyze:**

Guiding questions

How should I organize my data to perform analysis on it?

Fortunately, I had three separate data files that were simple, so the organization was not an issue.

Has my data been properly formatted?

Yes, Date / Decimal

What surprises did I discover in the data?

The current rate of gas is comparative to our all-time high and the likelihood of the it is surpassing is very great.

What trends or relationships did I find in the data?

It is increasing rapidly and has an exceptionally good chance of breaking all-time records.

How will these insights help answer theses business questions?

It can determine if the production cost / expenses is worth processing at the current gas rate or if it is better off waiting for.

Key tasks

1. Aggregate the data so it’s useful and accessible.

2. Organize and format the data.

3. Perform calculations.

4. Identify trends and relationships.

\*Deliverable \*

A summary of my analysis

**Share:**

Now that I have performed my analysis and gained some insights into the data, create visualizations to share my findings. The project manager has reminded me that they should be sophisticated and polished in order to effectively communicate to the executive team.

Guiding questions

Was I able to answer the business question?

Yes

What story does my data tell?

The data show that we are on the brink of surpassing record breaking numbers for the U.S average gallon of gas pricing.

How do I findings relate to my original question?

It breaks down the costs and can lead companies to decide if the cost / expense is something they can handle or if it may be something they need to re-analyze.

Who is my audience?

The client and boss that I am presenting my findings to for the average price per gallon of gas increase.

What is the best way to communicate with them?

I think it would be best to show my visualization to them to display how gasoline prices may impact all aspects of their business.

Can data visualization help me share my findings?

Yes, very well.

Is my presentation accessible to my audience?

Yes, I shared my finding with LinkedIn members that were posting about #gasprices. They were easily able to recognize the data and gave vital feedback. It would only make sense that my audience would be able to understand since they are experts in this field.

Key tasks:

1. Determine the best way to share my findings.

2. Create effective data visualizations.

3. Present my findings.

4. Ensure my work is accessible.

\*Deliverable\*

Supporting visualizations and key findings

**Act:**

Now that I have finished creating my visualizations, act on my findings. Organize the deliverables I created, including my top high-level insights based on the analysis.

Guiding questions:

What is my final conclusion based on the analysis?

How could my team and business apply my insights?

What next steps would I or my stakeholders take based on my findings?

Is there additional data I could use to expand on my findings?

Key tasks:

1. Create my portfolio.

2. Add my case study.

3. Practice presenting my case study to a friend or family member.

\*Deliverable\*

My top three recommendations based on my analysis