**Nick Vecchio**

**Final Project Report**

**Driving Question and Data Approach**

**Question:** What should parents invest in to secure their children's financial future?  
**Approach:**  
To address this, we analyzed historical and current data on gold prices, inflation rates, and stock market performance. These datasets were combined and cleaned to create an interactive data visualization dashboard that allows users to compare and analyze trends over time.

**Data Sources:**

1. Gold Prices:
   * Historical: National Mining Association (1833 to ~1980).
   * Recent: World Gold Council (1980 to present).
   * Combined dataset: DataHub’s aggregated gold price data.
2. Inflation Rates:
   * Bureau of Labor Statistics, monthly inflation rates (processed into a normalized format).
3. Stock Market Data:
   * S&P 500 historical prices from Investing.com.

**Data Sources, Quality, and Structure**

**Data Sources:**

* Gold Price data was sourced from aggregated historical datasets. One question I was unable answer is if my merging of data sets corrupted values enough to be inaccurate. In my initial investigation that did not appear to be the case.
* Inflation Data: Cleaned and normalized(to gold data format) to include monthly values.
* Stock Data: Filtered and normalized to match gold and inflation date ranges, ensuring consistency.

**Data Quality Issues:**

* Missing values in inflation data which were thankfully at the end of the data set and were dropped.
* Discrepancies in overlapping date ranges for gold prices were resolved by prioritizing combined dataset accuracy.
* Filtering date ranges (e.g., post-1971 gold standard removal) ensured that information was useful for current times and how it relates to inflation.

**Structure:**  
Each dataset was transformed into a uniform time-series format. Gold and stock prices were normalized for direct comparability, and inflation rates were similarly normalized.

**Data Structure**  
All datasets were cleaned, tidied, and normalized to align on a common time frame, starting in February 1970. This ensured consistency for visual comparisons and calculations, such as annualized growth rates.

**3. Data Wrangling, Analysis, and Visualization Development**

**Data Wrangling**

* Missing values in the inflation dataset were addressed by dropping them. They were values not yet in existence because that time had not happened yet.
* Dates were reformatted to YYYY-MM for consistency vs the gold data set.
* Price data was cleaned to remove invalid characters (e.g., commas).

**Analysis**

* Calculated annualized growth rates using logarithmic growth models to evaluate investment performance.
* Normalized datasets to facilitate direct comparisons between gold, stocks, and inflation.

**Visualization Development**

* **Tools Used:** Python, Dash, and Plotly, Numpy.
* **Visualizations Included:**
  + Line charts to show long-term trends in gold and stock prices.
  + Interactive gauges to visualize annualized growth rates.
  + Stacked bar charts comparing annual performance.

**4. Application of Course Concepts**

* **Data Organization:**  
  Tidy data principles ensured a consistent and analyzable structure across all datasets. Had to transform wide data (inflation, stock price) into a long format matching the gold data set.
* **Interactivity:**  
  This is one are where I wish I had the time to do more. Currently, many of the visuals use the standard interactions. I did create a play button for my gauge animation. Additionally, I had a visual working that would allow a user to see yearly growth of stock vs gold and then click into the year to see that years monthly growth. This was to show that there is volatility in the market. That did not make it into the Dash app yet.
* **Analysis Techniques:**  
  Compound growth calculations, normalization, and statistical comparisons were used.

**5. Audience-Centered Design**

The audience for this project is parents seeking straightforward insights into investment options. Design decisions prioritized simplicity and clarity:

* Used gauges and bar charts for quick comparisons.
* Created a user-friendly dashboard with clearly labeled buttons and cards.
* Simplified technical concepts for a non-expert audience.

Future improvements could include a dedicated learning page to explain key concepts like inflation and annualized growth rates.

**6. Limitations and Next Steps**

**Limitations**

* Data cleaning was time-intensive, giving me less time for advanced interactivity and polishing the dashboard.
* Incomplete data on external factors like wars, economic crises, and central bank gold purchases not only limited usefulness of data but also limited the interactions I had planned.

**Next Steps**

* Use API’s for up-to-date live information.
* Visual to show how gold price is affected by national banks (i.e.-Federal Reserve Bank) buying gold bullion.
* Events checkbox to show annotations for things like political change, wars/conflicts, economic crises, etc.
* Redesigned layout and style to be a “website” data dashboard.
* Definitions and Learning page.
* Increased interactivity in the data visualizations.
* Average cost of goods in USD from BLS

**Reflections**

I feel like the initial steps are the hardest and most critical to get right. Which includes sourcing your data, cleaning and tidying it and then normalizing it for display with your other data sets. This step would prove to be the most difficult and caused me to have the most sleepless nights. I am hopeful that with time and practice this step will get easier.

How I plan to use this knowledge (of data viz programming):

One of the big things I am passionate about is data democratization. Which seeks to make data available to everyone and easy to understand for all audiences.

This is something that is discussed all the time at KSU. How can we gain insights from all this disparate data? I would LOVE to create a dashboard for even one of our functional units to allow them to understand their data better and then use that information to make better decisions. Looking back, I do wish I had done my higher-ed focused project idea. However, I believe this step was important. Creating something for a topic I am passionate about kept me going. Because honestly there were many nights where I had already worked (programmed) all day for my job and then had to stay in the same chair programming all night for this class.

I love programming but doing it for 12 hours a day most days on top of family responsibilities, sickness, childcare issues, my physical disabilities and pain, was extremely difficult.

There were so many things I wish I had done in retrospect and would have done had I had more time. As I mentioned earlier, I just have so little time, that I am a little disappointed in the final product. Yes, I am glad it is working. But, I know I am capable of so much more. Usually, the things I create are highly polished. I had to give myself some slack though because considering the circumstances, the fact that I have a working, and not completely awful looking dashboard is a big accomplishment. I also have an extreme desire to further develop it.

Truly, I enjoyed this project and was happy learning new things. I cannot wait to develop it further and then take that knowledge and apply it to my job. Data wrangling and dashboarding are two skills in high demand currently at KSU.