**Q1.**

Five steps to cleaner data are:

1. Develop a data quality plan
2. Correct data at the source
3. Measure data accuracy
4. Manage data and duplicates
5. Append data

Six best practices

Insufficient understanding of the cause of anomalies makes it challenging to make the appropriate changes

Data deletion is the process of removing information from a system such that incomplete data cannot be precisely "filled in."

Continuous upkeep can be costly and time-consuming.

Building a data purification graph in advance to help with the process is challenging.

Think about your data holistically, taking into account not just who will be performing the analysis but also who will be utilising the findings.

Tighter controls over database inputs can guarantee that the system uses cleaner data.

Select software programmes that can identify errors in data and maybe fix them before they cause problems.

**Q2.**

Yes, I agree that reliable and accurate insights can only be produced by using high-quality datasets, and that data cleansing is a crucial step in data analysis and data science projects. However, the specific techniques and tools used for data cleaning may vary, and the choice of tools should be determined by the characteristics of the dataset, the nature of the data quality issues, and the goals of the study.

It is imperative to assess the particular needs of each project and choose data cleaning techniques and materials accordingly. While OpenRefine and Trifacta Wrangler are the main subject of this piece, you can use a variety of alternative data cleaning tools and techniques; the one you select will depend on the particulars of your project and the kinds of data anomalies that need to be corrected.

**Q3.**

Sources of errors not provided by author are:

Abnormalities and Exceptions: If anomalies or outliers are not adequately discovered and managed, they can cause errors and distort analyses.

Data Format modifications: If the structure or schema of the data source is altered, mistakes may arise that should be addressed properly, causing problems with data compatibility.

Mistakes in migration of data: Migration mistakes can happen during the transfer of data across databases or systems, leading to misaligned or corrupted data.

Incompatible Quantities of Measure: When working with numerical data, variations in the units of measurement might cause mistakes in computations and comparisons.

Breach of Security of data: Unauthorised modifications or corruption of data might result from data breaches, which need to be fixed throughout the data cleaning procedure.

**Q4.**

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

**References and Citations:**

 Dilmegani, C. (2023, October 12). Guide to Data Cleaning in &#039;23: Steps to Clean Data &amp; Best Tools. AIMultiple. https://research.aimultiple.com/data-cleaning/

In-Text Citation: (Dilmegani, 2023)

Hooland, S. V., Verborgh, R., & Wilde, M. D. (2019, April 10). Nettoyer ses données avec OpenRefine (S. Clochet, Trans.). Programming Historian En Français, 1. https://doi.org/10.46430/phfr0004

In-Text Citation: (Hooland et al., 2019)

McFarland, A. (2022, April 28). 10 Best Data Cleaning Tools. Unite.AI. https://www.unite.ai/10-best-data-cleaning-tools/

In-Text Citation: (McFarland, 2022)