**Analyzing Retail Sales and Customer Behavior**

Deeksha Chawla

Bellevue University

DSC 540 Data Preparation

Prof. Catherine Williams

May 28th, 2024

**Project Subject Area:**

Analyzing Retail Sales and Customer Behavior

**Changes to the Data:**

* Data from CSV and HTML sources were loaded, cleaned, and integrated into a single dataset.
* Weather data was fetched from an external API and added to the integrated dataset, enriching it with additional information.
* Normalization of keys and renaming of columns were performed to ensure consistency and ease of data manipulation.
* The final merged dataset underwent transformations such as formatting, filtering, and merging to prepare it for analysis and visualization.

**Legal or Regulatory Guidelines:**

* Depending on the specific context of the data and its usage, there may be legal and regulatory guidelines related to data privacy, intellectual property rights, and consumer protection laws.
* For example, handling customer data (such as city names) requires compliance with data protection regulations like GDPR in the EU or CCPA in California.

**Risks of Transformations:**

* Risks could include introducing biases or inaccuracies during data cleaning and transformation processes.
* For instance, improper handling of missing values or erroneous data could lead to incorrect analysis and decision-making.
* Additionally, integrating external data sources (such as weather data) may introduce inconsistencies or errors if not validated properly.

**Assumptions in Cleaning/Transforming Data:**

* Assumptions might have been made regarding the accuracy and completeness of the data from the CSV, HTML, and weather API sources.
* Assumptions may also have been made during data cleaning steps, such as the method used to handle missing values or outliers.

**Data Sourcing and Verification:**

* Data from the CSV and HTML sources were obtained from known and reputable sources, such as Amazon's bestsellers list.
* Weather data was sourced from the OpenWeatherMap API, which is a widely used and reliable source for weather information.
* Verification of the credibility of the data involved cross-referencing information from multiple sources and ensuring consistency and reliability.

**Ethical Acquisition of Data:**

* Data acquisition involved obtaining consent or using publicly available data sources.
* Care was taken to ensure that personal or sensitive information was handled appropriately and in compliance with ethical standards and legal requirements.

**Mitigation of Ethical Implications:**

* To mitigate ethical implications, transparency in data handling processes was maintained through documentation and clear communication.
* Privacy concerns were addressed by anonymizing or aggregating sensitive information where necessary.
* Compliance with relevant regulations and guidelines was ensured throughout the data handling and analysis process.
* Regular audits and reviews of data handling practices could be conducted to identify and address any potential ethical concerns proactively.