**Teddy Aigbedion**

**Acquiring and Wrangling Data Assessment 1A**

**1. Point of Sale (POS) Data**

**Cost**

* **High:** Maintaining and updating a relational database with daily transactional data can be costly due to storage, maintenance, and infrastructure requirements.

**Accuracy**

* **High:** Internal records are typically accurate, especially if the data is directly captured from sales transactions without manual entry.

**Completeness**

* **High:** POS data is likely to be complete as it includes all transactions from both physical and online sales.

**Alignment with other data sources or questions**

* **Good:** Aligns well with Customer Surveys and Website Analytics as it can provide detailed transaction data to correlate with customer feedback and online behavior.

**Inclusion/exclusion criteria**

* **Clear:** Includes transactional details; may exclude non-transactional interactions like browsing without purchase.

**Compliance with legal and ethical standards**

* **Good:** limiting access to authorized personnel helps ensure compliance with privacy laws.

**Format/location/collection method**

* **Structured data in a relational database:** They are usually located in databases e.g. Microsoft Sql Server, Oracle Postgrel etc. SQL and other programming language like Python, Java are used in collecting or querying data

**2. Customer Surveys**

**Cost**

* **Moderate:** Periodic surveys incur costs related to survey design, distribution, and data processing, but are less frequent than continuous data collection.

**Accuracy**

* **Variable:** Depends on the honesty and accuracy of the respondents. Bias and response errors can affect data quality.

**Completeness**

* **Moderate:** May not capture every customer's opinion, leading to potential gaps in the dataset.

**Alignment with other data sources or questions**

* **Good:** Can complement POS Data by providing context to transactional data (e.g., reasons behind purchase decisions).

**Inclusion/exclusion criteria**

* **Clear:** Includes feedback from surveyed customers, excluding those who did not participate.

**Compliance with legal and ethical standards**

* **Good:** Surveys should be conducted with explicit consent and respect for privacy.

**Format/location/collection method**

* **Structured data in a CSV file:** Simple to handle but may require conversion for integration with other data formats. Collection method usually include forms filling, both physical and online

**3. Website Analytics**

**Cost**

* **Moderate to High:** Depending on the tools used, real-time analytics and storage of detailed user behavior data can be costly.

**Accuracy**

* **High:** Generally accurate as it is automatically captured by analytics tools, though it may depend on the correct setup of tracking mechanisms.

**Completeness**

* **Moderate:** Captures detailed user interactions but may miss data if users block tracking or if there are tracking errors.

**Alignment with other data sources or questions**

* **Good:** Provides valuable insights into user behavior that can be linked to POS Data and Customer Surveys for a comprehensive view of the customer journey.

**Inclusion/exclusion criteria**

* **Clear:** Includes online interactions; excludes offline behavior unless integrated with POS Data.

**Compliance with legal and ethical standards**

* **Moderate:** Must comply with data privacy regulations like GDPR and CCPA, requiring proper consent and data anonymization.

**Format/location/collection method**

* **Structured data in JSON format:** Good for detailed data, and usually require transformation for use with other data source, and method of collection includes Python modules like BeautifulSoup, Pandas etc

**4. Social Media Monitoring**

**Cost**

* **High:** Continuous monitoring and data storage can be expensive due to API costs and data volume.

**Accuracy**

* **Variable:** Social media data can be noisy and contain irrelevant information; sentiment analysis accuracy depends on the algorithms used.

**Completeness**

* **Moderate to Low:** May not capture all mentions due to privacy settings, platform restrictions, and data bloat policies.

**Alignment with other data sources or questions**

* **Good:** Complements POS Data and Customer Surveys by providing real-time feedback and sentiment analysis from social media platforms.

**Inclusion/exclusion criteria**

* **Clear:** Includes mentions and posts about the company; excludes private messages and data not accessible via APIs.

**Compliance with legal and ethical standards**

* **Moderate to Low:** Must ensure compliance with platform-specific policies and regulations on data collection and usage.

**Format/location/collection method**

* **Unstructured text data in JSON format:** Requires significant processing and analysis to extract meaningful insights and align with structured data sources, method of collection usually is python modules like BeautifulSoup.

**Summary**

* **POS Data:** High cost and accuracy, highly complete, well-aligned with other sources, clear inclusion criteria, good legal compliance, structured format.
* **Customer Surveys:** Moderate cost, variable accuracy, moderate completeness, well-aligned with other sources, clear inclusion criteria, good legal compliance, structured format.
* **Website Analytics:** Moderate to high cost, high accuracy, moderate completeness, well-aligned with other sources, clear inclusion criteria, moderate legal compliance, structured format.
* **Social Media Monitoring:** High cost, variable accuracy, moderate to low completeness, well-aligned with other sources, clear inclusion criteria, moderate to low legal compliance, unstructured format.

**Each dataset has its strengths and challenges, and using them in combination can provide a comprehensive view of customer behavior and feedback. Ensuring proper data integration and maintaining legal and ethical standards are key to leveraging these datasets effectively.**

**PROPOSAL RESPONSE**

### Proposal 1: Socialytics Marketing

#### Cost

* **Initial Cost:** $12,000 for historical analysis
* **Ongoing Cost:** $20,000 annually for ongoing monitoring, analysis, and reporting

#### Accuracy

* **High:** Uses state-of-the-art tools for real-time social media monitoring and sentiment analysis. However, accuracy depends on the quality and algorithms of the tools used.

#### Completeness

* **Moderate to High:** Provides comprehensive social media insights but might miss offline consumer interactions. Completeness is ensured through continuous monitoring.

#### Alignment with other data sources or questions

* **Good:** Aligns well with other marketing data sources such as POS Data and Customer Surveys, providing a broad view of consumer sentiment and engagement.

#### Inclusion/exclusion criteria

* **Clear:** Includes all relevant social media data; excludes non-social media interactions.

#### Compliance with legal and ethical standards

* **Good:** Compliance with social media platform policies and privacy regulations is essential. Assumes adherence to these standards.

#### Format/location/collection method

* **Real-time monitoring:** Uses advanced tools to collect and analyze unstructured social media data in JSON format.

### Proposal 2: Consumer Insights

#### Cost

* **Initial Cost:** $50,000 for the project covering all research activities, analysis, and reporting.

#### Accuracy

* **High:** Specializes in consumer behavior analysis and market trends with a strong focus on survey-driven data and quarterly sales reports.

#### Completeness

* **High:** Provides a comprehensive understanding of consumer behavior and market trends, ensuring detailed insights.

#### Alignment with other data sources or questions

* **Good:** Complements POS Data and Customer Surveys by providing in-depth analysis of consumer behavior and market trends.

#### Inclusion/exclusion criteria

* **Clear:** Includes consumer behavior patterns, preferences, and purchasing habits; excludes non-consumer-related data.

#### Compliance with legal and ethical standards

* **High:** Strong focus on compliance with privacy and ethical standards in consumer research.

#### Format/location/collection method

* **Structured data:** Survey-driven data in CSV format, quarterly sales reports, and structured reports.

### Proposal 3: Maritime Insights Research Solutions

#### Cost

* **Initial Cost:** $15,000 annual access to existing curated market data.

#### Accuracy

* **High:** Known for providing highly accurate and timely data for predictive pricing in maritime industries.

#### Completeness

* **Moderate:** Focuses on maritime industry data, which may not be directly relevant to CPFC Industries’ retail and consumer market focus.

#### Alignment with other data sources or questions

* **Low to Moderate:** Less alignment with CPFC’s consumer-focused data needs, as it specializes in maritime industry data.

#### Inclusion/exclusion criteria

* **Clear:** Includes market pricing, fishing routes, weather, etc.; excludes non-maritime data.

#### Compliance with legal and ethical standards

* **Good:** Assumes adherence to legal and ethical standards for data collection and usage.

#### Format/location/collection method

* **Curated data sets:** Structured data in formats suitable for predictive analysis in the maritime industry.

### Summary

* **Socialytics Marketing:** Offers high accuracy and alignment with social media data needs, though it comes with a moderate to high cost. The completeness is good for online interactions but may lack offline insights.
* **Consumer Insights:** Provides the most comprehensive consumer behavior and market trends analysis at a higher initial cost. Completeness and alignment with CPFC's data needs are high, making it a strong option.
* **Maritime Insights Research Solutions:** While highly accurate and cost-effective for the maritime industry, its relevance to CPFC’s retail market needs is low.

### Recommendations

**For CPFC Industries, Consumer Insights seems to be the most aligned with their goals of understanding consumer behavior and market trends, despite the higher cost. Socialytics Marketing could be considered for a detailed view of social media interactions and engagement. Maritime Insights Research Solutions may not be suitable due to its focus on the maritime industry.**