**DATA COLLECTION PROCESS and TOOLS**

**Data Collection for Project Topic**

**Data Sources:**

Microsoft forms was used for data collection to interview MSMEs about the business's operations, challenges, and opportunities available to explore. The primary data collected through surveys and questionnaires for [SMEs](https://forms.office.com/Pages/ResponsePage.aspx?id=2wYhaHuaNUu5oIhHM4XWz_0NZd6Mt6xOvf715VWeQw5UMUwzRjcxNEdTNlhEUlhFOFRESlQzUEJYTy4u) and [blue-collar workers and artisans](https://forms.office.com/Pages/ResponsePage.aspx?id=2wYhaHuaNUu5oIhHM4XWz_0NZd6Mt6xOvf715VWeQw5UOU0yOTVZRzdDWFc4Wjc3RjkwRk1DVUxLSy4u) can be accessed.

**Data Extraction and Collection for Product Analysis**

**Data Sources:**

* The data was sourced as unstructured text and images stored in PDF documents from Small and Medium Enterprise Regulatory bodies in Nigeria and the National Bureau of statistics (SMEDAN, NBS).

**Data Extraction:**

* The text and numeric data were extracted using Jupyter notebook, MS Word, Adobe, and online document conversion websites into structured data stored in MS Excel tables following a predefined schema.

**Data Type:**

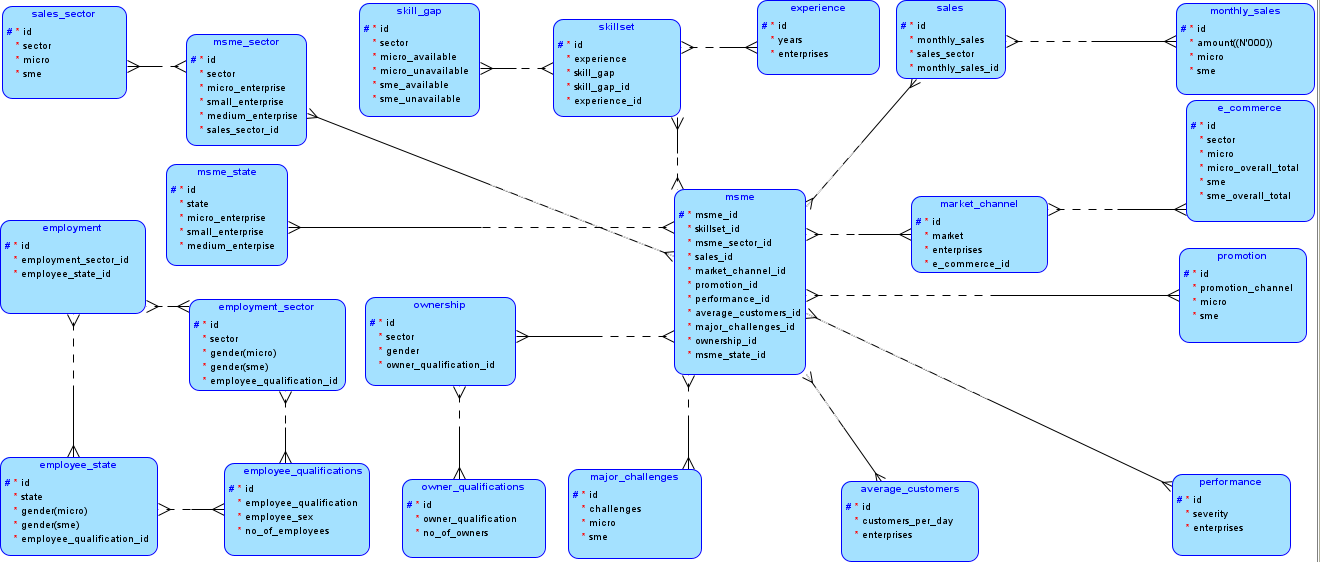
* Numeric data that quantifies the ecosystem of MSMEs
* Categorical data that provides contextual information about MSMEs.

**3.4 DATA ANALYSIS**

**Data Cleaning**

* The data was cleaned using Microsoft Excel and its functionalities such as formulas, tables, charts, macros. The data were modeled and analyzed to provide insights about the demographic and answer questions that explore these demographics.
* Macros was recorded for repetitive tasks such as text to column, changing text casing and computing the median value.
* Handling outliers. The Box plot was used to visualize data for skewness and Prescence of outliers. This means median was used to fill missing values where outliers were present and mean where none.
* Conditional formatting was used to tag outliers as they were included during analysis.
* Dealing with missing values. The center of tendencies, mean, median, mode was used when applicable.
* Winsorised mean was used instead in the case where the median was also affected by outliers.
* The data were stored as structured data in tables that were named.
* Duplicate values were handled by removal.
* The data was fortified with additional information where necessary, e.g., country, region and zip code of states.
* Spell checks were performed to ensure accuracy and consistency of text data.
* The data was preprocessed for analysis in Power BI by including primary and foreign keys to the dataset to show relationship.

**Database Design and Implementation**

A database was created to store the data extracted and collected for further future analysis.

**Database Documentation Process:**

Entity-Relationship Diagram (ERD) from modeling tool (Oracle modeler).

1. SQL schema script was written based off the ERD and entity relationship defined using primary and foreign keys. The data types were defined, and tables created for every entity.

Snippet:

/\* Database schema to keep the structure of entire database. \*/

CREATE TABLE msme(

    id INT GENERATED ALWAYS AS IDENTITY PRIMARY KEY,

    skillset\_id NUMERIC,

    sales\_id NUMERIC,

    market\_channel\_id NUMERIC,

    promotion\_id NUMERIC,

    performance\_id NUMERIC,

    average\_customers\_id NUMERIC,

    major\_challenges\_id NUMERIC,

    ownership\_id NUMERIC,

    msme\_state\_id NUMERIC

);

CREATE TABLE msme\_state(

    id INT GENERATED ALWAYS AS IDENTITY PRIMARY KEY,

    state VARCHAR,

    micro\_enterprise INT,

    small\_enterprise INT,

    medium\_enterprise INT

);

1. SQL script for populating the database and creating tables were created using individual .CSV files. The macros enable worksheets were converted to .CSV files and inserted into the database.

Code snippet:

/\* Populate database with sample data. \*/

COPY msme (skillset\_id, sales\_id, market\_channel\_id, promotion\_id, performance\_id, average\_customers\_id, major\_challenges\_id, ownership\_id, msme\_state\_id) FROM '/path/to/msme.csv' DELIMITER ',' CSV HEADER SKIP 1;

COPY msme\_state (state, micro\_enterprise, small\_enterprise, medium\_enterprise) FROM 'C:/Users/DELL/Desktop/HANDY/msme\_state.csv' DELIMITER ',' CSV HEADER SKIP 1;

COPY msme\_sector (sector, micro\_enterprise, small\_enterprise, medium\_enterprise, sales\_sector\_id) FROM 'C:/Users/DELL/Desktop/HANDY/msme\_sector.csv' DELIMITER ',' CSV HEADER SKIP 1;

1. SQL Queries for Data Verification were written to be implemented. Comments are provided for adequate documentation.

Snippet:

/\*Queries that provide answers to the questions for data verification.\*/

-- Check for NULL values in important columns

SELECT \*

    FROM msme

    WHERE skillset\_id

    IS NULL OR sales\_id IS NULL OR msme\_state\_id IS NULL;

-- Check for duplicate entries

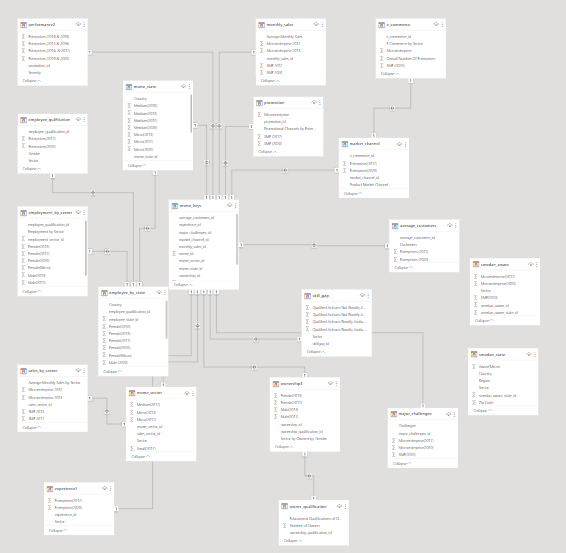
SELECT \*, COUNT(\*) FROM msme

    GROUP BY skillset\_id, sales\_id, market\_channel\_id, promotion\_id, performance\_id, average\_customers\_id, major\_challenges\_id, ownership\_id, msme\_state\_id

    HAVING COUNT(\*) > 1;

**Data Analysis**

* Descriptive and exploratory data analysis was performed using descriptive statistics, charts, and graphs, and reported using tools such as Power BI, MS Excel, Python EDA, and visualization libraries (seaborn, matplotlib, pandas)**.**
* A schema was created to define the relationship between tables.

****

* Metrics were identified such as number of enterprises by categories and average monthly sales turnover.
* Design principles were implemented in dashboard design.
* Visualisation: The charts were carefully chosen in consideration of data types and communicating insights effectively.

**Metrics to Measure**

1. Number of MSMEs available.
2. Average monthly sales turnover.
3. Educational qualifications
4. Percentage of gendered distribution of owners and employees.

4.0   RESULT

**User Research**

To ensure the project's relevance and real-world impact, Team Fem-tastic conducted extensive research to understand the needs of users seeking skilled services in the informal labor market. The research involved engaging with artisans and potential users who require the services of artisans.

**Objective of the Analysis:**

**Hypothesis**

From questionnaires distributed online

**H0:** An online talent marketplace for artisans would not significantly improve the quality of services for microenterprises and ease the hiring process and quality of artisans hired for SMEs.

**H1:** An online talent marketplace for artisans would not significantly improve the quality of services for microenterprises and ease the hiring process and quality of artisans hired for SMEs.

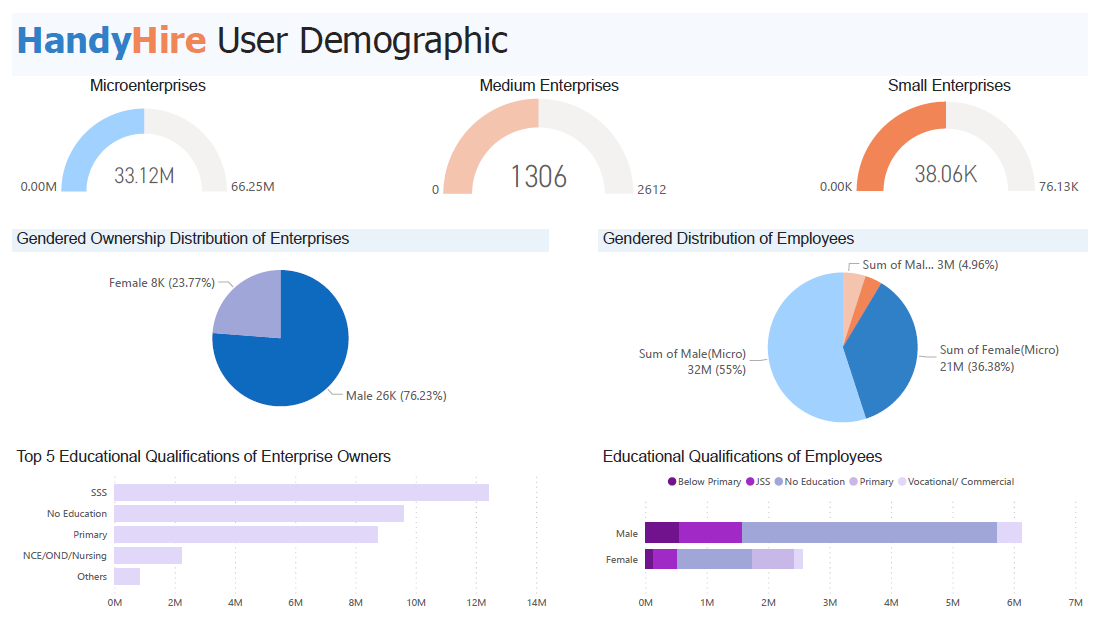
**Questions**

The data analysis aims to answer these questions:

1. What metrics can be measured to pinpoint the success of HandyHire?
2. Who is the user demographic for HandyHire?
3. What is the average monthly sales turnover for MSME?
4. What are the markets these enterprises already operate in?
5. What are the challenges faced by these enterprises that can be alleviated by HandyHire?
6. What are the communication and promotional channels already used by and for these enterprises?

**Visualisation and Insights**

1. **Handy Hire User Demographic:** highlights the potential users of handy hire. It was found that microenterprises are readily available to be vetted for product and service quality at 33.1 million. SMEs across sectors are available at over 38 thousand SMEs. Employees and Owners are more likely to be Male than Female at 39.98% male to >36.38% female employers and 76.23% male to 23.73% female. This population is largely literate with non-tertiary and diploma level education (22 million) while approximately 14 million had no education.



1. Average Monthly Sales: More average monthly sales are recorded from enterprises by economic sector than they were within a sales range.

Microenterprises nationwide make N34,000,000 average monthly sales. Almost two times more than small and medium enterprises at N18,000,000.

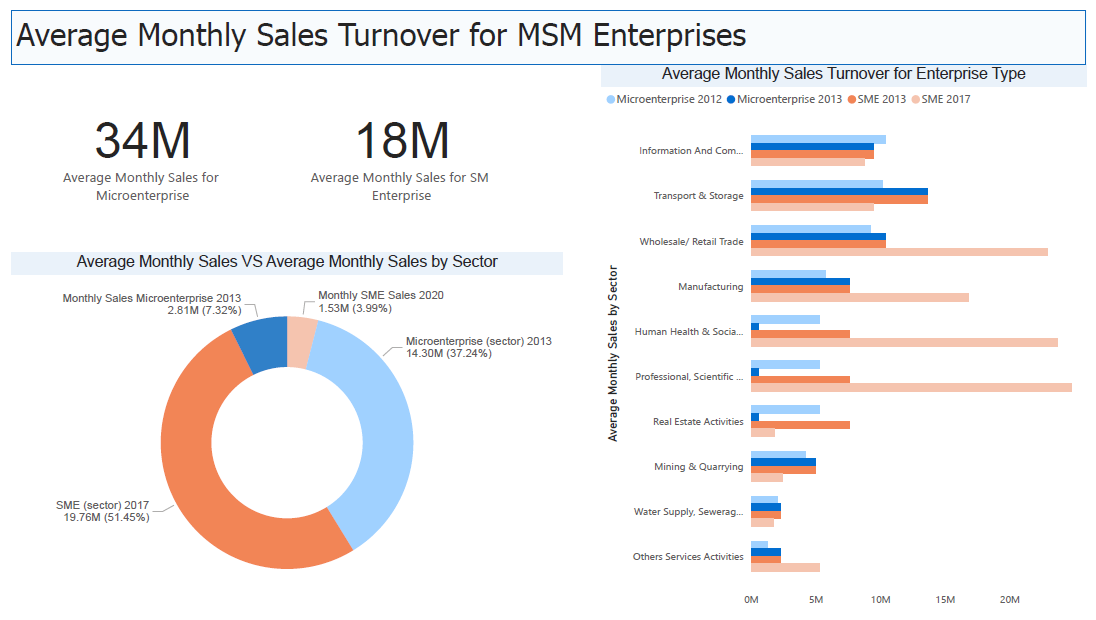
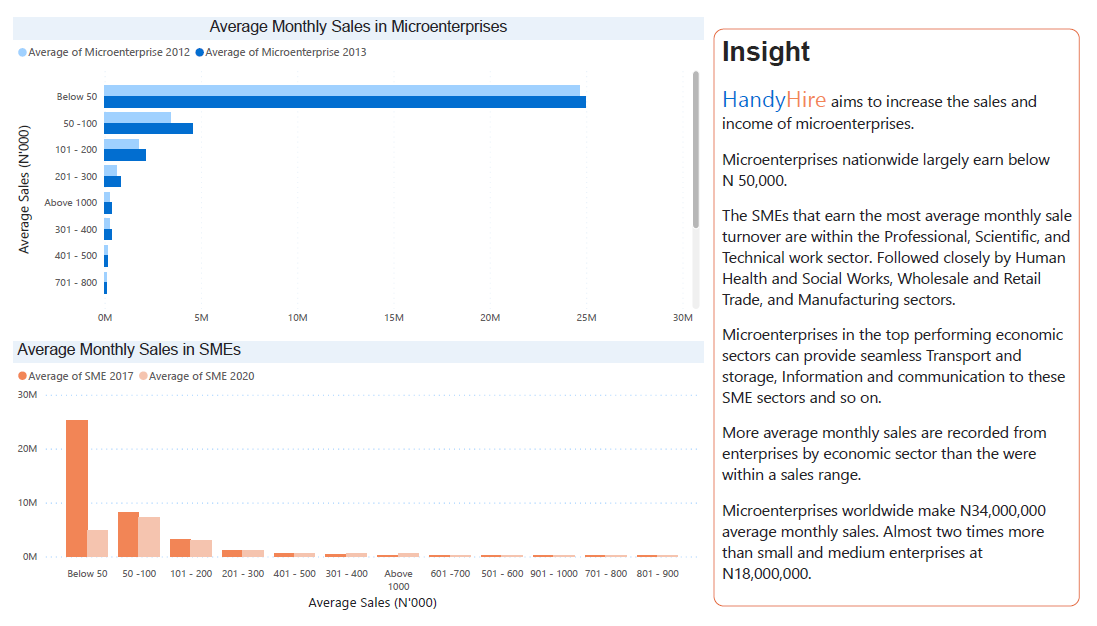
HandyHire aims to increase the sales and income of microenterprises.

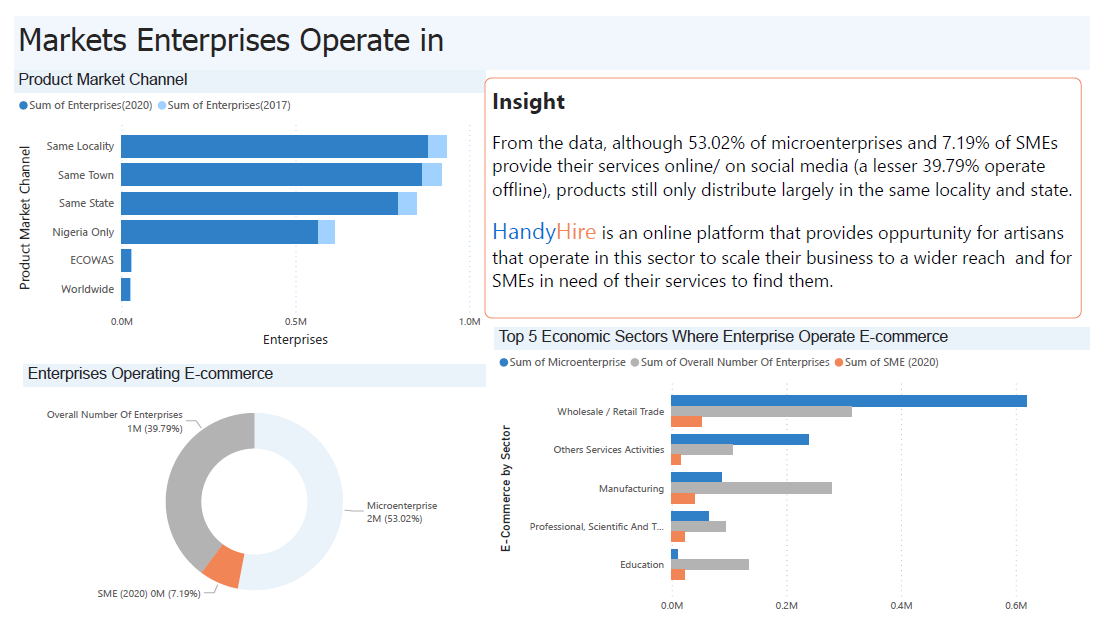
Microenterprises nationwide largely earn below N 50,000. This means an online marketplace that exposes the goods and services of these microenterprises can increase their average monthly sales.

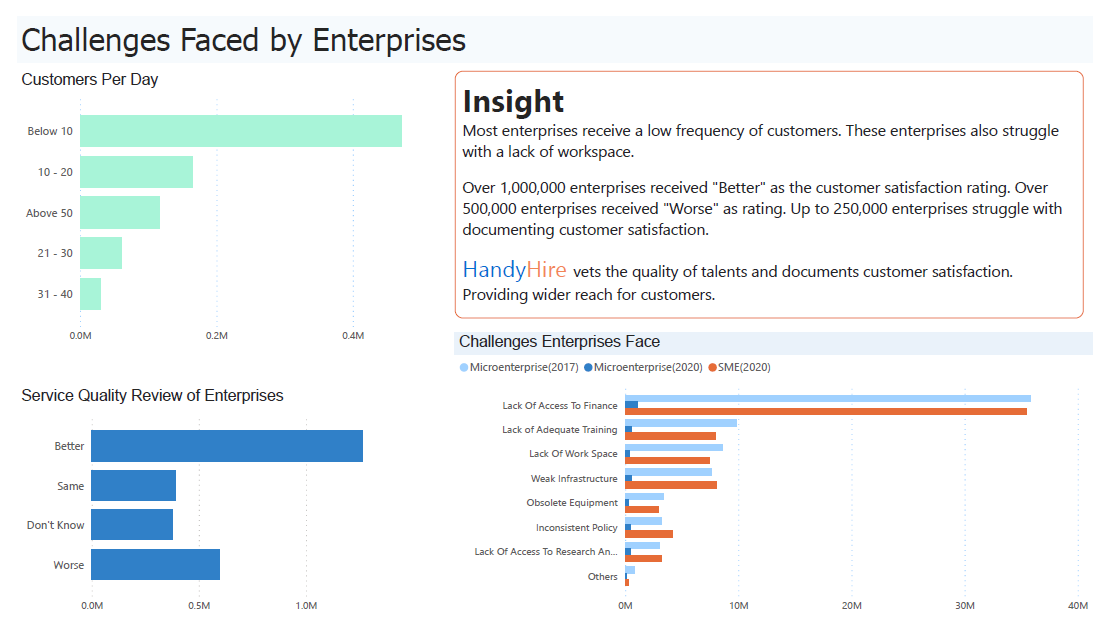
The SMEs that earn the most average monthly sale turnover are within the Professional, Scientific, and

Technical work sector. Followed closely by Human Health and Social Works, Wholesale and Retail Trade, and Manufacturing sectors. The Artisans that operate in this economic sector can be beneficiaries.

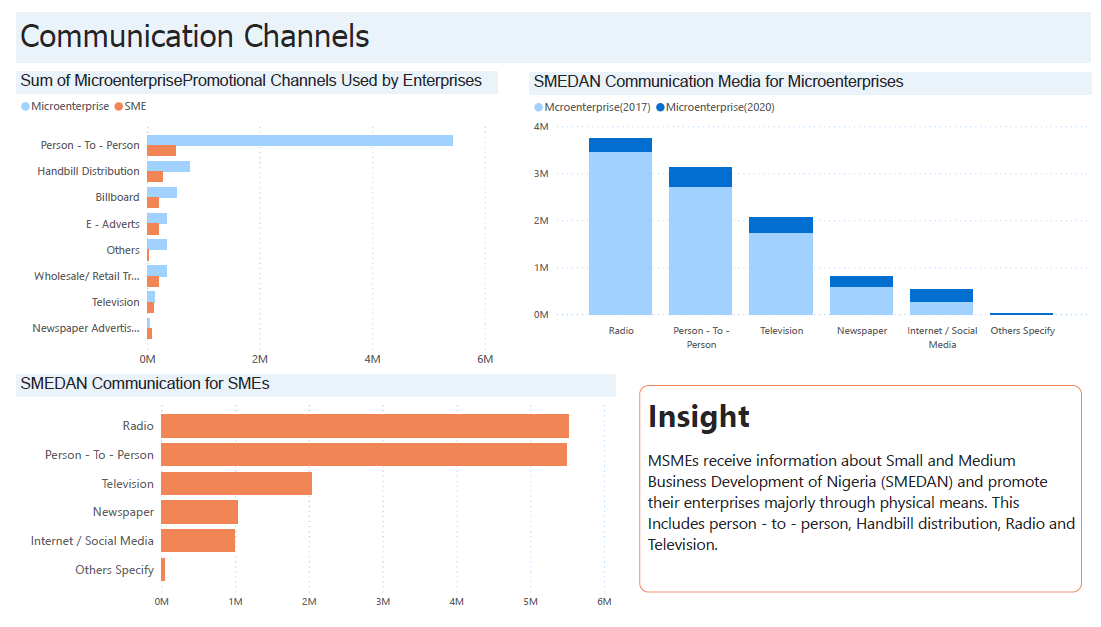
Microenterprises in the top performing economic sectors can provide Transport and storage, Information, and communication to these SME sectors and so



1. **Market channel:** explores the markets open to enterprises and provides insight into existing e-commerce ventures.



1. **Challenges:** Explores the challenges these enterprises face and provides insights.

E. **Communication Channels:** Explores communication channels readily available to enterprises.

**Report**

The raw data, Macro Enabled workbook clean and manipulated data, CSV, ERD and SQL queries, and Power BI desktop document have been saved to Google Drive and can be accessed [here.](https://drive.google.com/drive/u/2/folders/1f3P5ATGiHtHmUPn4PGPrbXaNVVuqha2g) Please download the workbook to enable macros functionalities.

**References:**

* 1. Statiista, Apr 8, 2022, *Number of micro, small, and medium-sized enterprises (MSMEs) in Nigeria 2017*, visited April 25, 2024, <https://www.statista.com/statistics/1300426/micro-small-and-medium-enterprises-number-in-nigeria/>
  2. Statista, Dec 13, 2022, *Number of people employed in micro enterprises in Nigeria 2020, by gender,* visited April 25, 2024, <https://www.statista.com/statistics/1300073/employment-in-micro-enterprises-in-nigeria-by-gender/>
  3. Statista, Dec 13, 2022, *Number of micro enterprises (MEs) in Nigeria 2017-2020, by sector*, visited April 25, 2024, < <https://www.statista.com/statistics/1299780/number-of-informal-micro-enterprises-in-nigeria-by-sector/>>
  4. Small and Medium Enterprises Development Agency of Nigeria (SMEDAN), National Bureau of Statistics (NBS), “*SMEDAN AND NATIONAL BUREAU OF STATISTICS COLLABORATIVE SURVEY: SELECTED FINDINGS (2013)”,* 2013.< [www.nigerianstat.gov.ng](http://www.nigerianstat.gov.ng)>
  5. SMEDAN & NBS, *“National Survey of Micro Small and Medium Enterprises(MSMEs)”*, 2017, < <https://www.google.com/url?sa=t&source=web&rct=j&opi=89978449&url=https%3A%2F%2Fnigerianstat.gov.ng%2Fpdfuploads%2FSMEDAN%2520REPORT%2520Launch%2520Presentation%25202017.pdf>>
  6. SMEDAN & NBS, “*2021 Survey MSME Report”, 2021. <* [*https://www.google.com/url?sa=t&source=web&rct=j&opi=89978449&url=https://nigerianstat.gov.ng/download/290*](https://www.google.com/url?sa=t&source=web&rct=j&opi=89978449&url=https://nigerianstat.gov.ng/download/290)*>*