**PYTHON PROJECT REPORT**

(Project Semester: January-April 2025)

**Title of the Project: Poverty By Income**

**Submitted by:**

**Name:Gunda Venkata Sai Gowtham**

**Registration No.: 12310658  
Programme and Section: B.Tech CSE (K23FA)  
Course Code: INT375**

**Under the Guidance of:  
Sandeep Kaur Mam(UID:23614)**

**Discipline of CSE/IT**  
**Lovely School of Computer Science & Engineering**  
**Lovely Professional University, Phagwara**

**DECLARATION**

I, **Gunda Venkata Sai Gowtham**, student of **Bachelors of Technology (B.Tech)** underCSE/IT Discipline at Lovely Professional University, Punjab, hereby declare that all theinformation furnished in this project report is based on my own intensive work and is genuine.

Date: 15-April-2025

Signature: Gunda Venkata Sai Gowtham  
Registration No.: 12310658  
Name of the Student: Gunda Venkata Sai Gowtham

# ****CERTIFICATE****

This is to certify that **Gunda Venkata Sai Gowtham** bearing Registration No. **12310658** has completed **INT375** project titled **“Poverty By Income”** under my guidance and supervision. To the best of my knowledge, the present work is the result of her original development, effort, and study.

**Sandeep Kaur Mam  
School of Computer Science & Engineering**

**Lovely Professional University**  
**Phagwara, Punjab**

Date: **15-April-2025**

**ACKNOWLEDGMENT**

I would like to express my sincere gratitude to **Sandeep Kaur Mam**, my project guide, for their invaluable support, guidance, and encouragement throughout the development of this project. Their expert insights and constructive feedback have been instrumental in shaping the project's outcome.

I am also thankful to **Lovely Professional University** for providing a conducive learning environment and access to resources that made this project possible. Additionally, I extend my appreciation to my professors and peers for their continuous motivation and insightful discussions, which greatly enhanced my understanding of the subject.

Lastly, I would like to acknowledge the unwavering support of my family and friends, whose encouragement has been a source of inspiration throughout this journey.

# ****TABLE OF CONTENTS****

1. [Introduction](#_1._INTRODUCTION)
2. [Source of Dataset](#_2._SOURCE_OF)
3. [Dataset Preprocessing](#_3._DATASET_PREPROCESSING)
4. [Analysis on Dataset (for each objective)](#_4._ANALYSIS_ON)
   * [i. General Description](#_i._General_Description)
   * [ii. Specific Requirements](#_ii._Specific_Requirements)
   * [iii. Analysis Results](#_iii._Analysis_Results)
   * [iv. Visualization](#_iv._Visualization)
5. [Conclusion](#_5._CONCLUSION)
6. [Future Scope](#_6._FUTURE_SCOPE)
7. [References](#_7.REFERENCES)

# ****1. INTRODUCTION****

# In the era of evidence-based policymaking and global development efforts, understanding the distribution and trends of poverty is essential for designing effective social and economic interventions. One of the most powerful tools in this effort is Exploratory Data Analysis (EDA)—a process that involves summarizing the main characteristics of a dataset through visual and statistical techniques. EDA helps uncover hidden patterns, identify data quality issues, test hypotheses, and highlight key trends that may inform decision-making at both national and global levels.

# This project, titled “Global Poverty Analysis Using Python”, employs EDA to examine poverty levels across various countries and years using real-world data. The dataset includes the percentage and number of people living under different income thresholds (e.g., $2.15/day, $6.85/day, and $10/day), providing a detailed picture of poverty distribution worldwide. The goal of this project is to analyze poverty indicators, visualize trends over time, and compare poverty levels across countries using Python’s data science stack.

# Python, along with its robust libraries like Pandas, Matplotlib, and Seaborn, serves as the backbone of this analysis. These tools allow us to efficiently process, manipulate, and visualize complex data to gain meaningful insights.

# Key goals of this project include:

# Reading and preprocessing global poverty data.

# Performing statistical analysis on poverty indicators across different countries and years.

# Identifying countries with the highest or lowest poverty rates at specific income thresholds.

# Visualizing temporal trends and cross-country comparisons through charts and graphs.

# Supporting research, policymaking, and educational objectives through intuitive and insightful visualizations.

# Unlike traditional systems that merely store and display statistics, this project focuses on deep exploration of the data to uncover stories behind the numbers. From identifying which countries have made progress in reducing extreme poverty to visualizing income distribution patterns, EDA transforms raw data into actionable insights.

# This project is especially useful for:

# Researchers and policymakers studying global poverty and inequality.

# Students and educators learning how to apply data science techniques in the context of international development.

# Data analysts and visualization experts who want to build compelling visual narratives using real-world socio-economic data.

# In conclusion, this Global Poverty Analysis System is a compelling example of how Python and EDA can be combined to extract knowledge from data. By transforming numerical values into intuitive visuals, it empowers stakeholders to better understand and address the multifaceted challenge of poverty.

# ****2. SOURCE OF DATASET****

# The dataset used for this analysis is titled “Poverty by Income” and is publicly available from open data repositories focused on global development statistics. It contains comprehensive information on the number and percentage of people living below various daily income thresholds across multiple countries and years.

# Key Features of the Dataset:

# Country: Nation for which poverty data is recorded

# Year: Year of data collection

# Poverty thresholds: Income levels ranging from below $1/day to $40/day

# Number of people below each threshold (absolute count)

# Percentage share of population living below each income level

# This dataset provides a valuable foundation for analyzing global poverty trends and assessing how income distribution has evolved over time across different regions.

# ****3. DATASET PREPROCESSING****

To ensure the dataset was suitable for analysis, a systematic data preprocessing phase was carried out. The raw dataset, sourced from a government data portal, contained monthly retail sales figures across various product categories. Upon loading the dataset, an initial review was conducted to understand the structure, format, and completeness of the data. This review revealed several inconsistencies and missing values that needed to be addressed.

The first step in preprocessing involved handling **missing data**. A detailed check was performed to identify any null or incomplete entries. Depending on the nature and significance of the missing values, different imputation techniques were used. For example, in time series columns, missing entries were filled based on previously observed values (forward fill) to maintain trend continuity. In numerical columns, mean or median values were used when appropriate to preserve the dataset's statistical balance. If certain rows or columns contained excessive missing data and did not contribute meaningfully to the analysis, they were removed.

Next, **data cleaning** was conducted. Redundant columns that did not offer analytical value were dropped. Column names were reformatted for consistency—removing special characters, converting to lowercase, and making names more readable. In cases where categorical data entries showed inconsistencies (e.g., varied naming for the same category), standardization was applied to unify them. This helped to avoid duplication and ensured that grouping and filtering operations would yield accurate results.

**Data type validation and conversion** formed another essential part of the preprocessing phase. Date fields were converted into a standard datetime format to support chronological sorting and time-based analysis. Numeric fields were checked to ensure all values were in the correct format and free of unexpected characters or text, which could interfere with computations. Ensuring correct data types allowed for smooth statistical operations and reliable visual representations.

To further enrich the dataset, **feature engineering** techniques were applied. New columns were created to support deeper analysis. For example, month and year were extracted from date entries to allow monthly trend analysis. In addition, sales differences over months, category contribution percentages, and cumulative figures were computed to bring out hidden patterns in the data. These derived features made it easier to compare and contrast performance across different retail categories and time periods.

Finally, the dataset was **sorted and filtered** to facilitate focused analysis. Categories were grouped based on sales volume, and the dataset was rearranged to highlight top-performing segments. Outlier detection was also performed to identify unusual spikes or drops in sales, which were examined further to understand their impact. Once the dataset was fully prepared, it was stored in a structured format, ready for visualization and exploratory data analysis.

This preprocessing phase was vital in transforming the raw dataset into a high-quality, analysis-ready format. It ensured that the data was not only accurate and complete but also tailored for meaningful insights and decision-making. Proper preprocessing greatly enhanced the reliability of the results obtained during subsequent stages of the project.

# ****4. ANALYSIS ON DATASET****

### **Objective 1: Compare the proportion of people living below the poverty line in selected countries for a specific year.**

**i. General Description:**A bar plot is used to compare poverty levels among countries for a specific year. This provides an at-a-glance understanding of which countries are most affected by extreme poverty.

**ii. Specific Requirements:**

* Data filtered for the year 2010
* Top 10 countries with the highest **% below $2.15/day**

**iii. Analysis Results:**  
In 2010, the top 10 countries exhibited significantly high poverty rates, with percentages exceeding 70% in some nations. These countries are mostly low-income and located in sub-Saharan Africa.

**iv. Visualization:**  
A horizontal bar plot clearly displays the top 10 countries with the highest poverty percentage, highlighting disparities between nations.

### **Objective 2: Explore the relationship between time and poverty rates for a single country.**

**i. General Description:**  
A scatter plot helps explore the change in poverty levels over time for a single country. This format is useful for detecting trends, fluctuations, or outliers.

**ii. Specific Requirements:**

* Country: **Albania**
* Indicator: **% below $2.15/day** over all available years

**iii. Analysis Results:**  
Albania shows a declining trend in extreme poverty over the years, indicating economic improvements and development programs' effectiveness.

**iv. Visualization:**  
The scatter plot, paired with a dashed trend line, visualizes the year-wise decrease in poverty levels for Albania.

### **Objective 3: Visualize and compare poverty trends over time between two or more countries.**

**i. General Description:**  
A line plot allows comparison of poverty trends across multiple countries. It's especially effective in visualizing longitudinal data.

**ii. Specific Requirements:**

* Countries: **Albania** and **Nigeria**
* Years: **2000 to 2020**
* Indicator: **% below $2.15/day**

**iii. Analysis Results:**  
Albania consistently shows lower poverty rates compared to Nigeria. While Albania's trend is downward, Nigeria exhibits fluctuations with persistently higher poverty levels.

**iv. Visualization:**  
A dual-line chart showing poverty trajectories for both countries helps in understanding how poverty is evolving in different socio-economic contexts.

### **Objective 4: Analyze the distribution and variability of poverty rates for a specific year**

### **i. General Description:** A box plot shows the distribution, median, quartiles, and outliers of poverty levels across all countries in a single year.

**ii. Specific Requirements:**

* Year: 2015
* Indicator: % below $6.85/day

**iii. Analysis Results:**The box plot reveals significant variability in poverty levels among countries. Outliers suggest countries with exceptionally high or low poverty rates in 2015.

**iv. Visualization:**The vertical box plot displays median poverty, interquartile range, and extreme values effectively.

### **Objective 5: Show the frequency distribution of poverty levels across countries in a given year.**

**i. General Description:**  
A histogram shows the frequency distribution of poverty rates across countries, indicating how values are spread.

**ii. Specific Requirements:**

1. Year: **2015**
2. Indicator: **% below $10/day**

**iii. Analysis Results:**  
Most countries have poverty rates between 20% and 60%, with fewer countries experiencing very low or very high rates.

**iv. Visualization:**  
The histogram, with bins dividing poverty rates into intervals, shows how common different poverty levels are across countries.

# 

# ****5. CONCLUSION****

# This project demonstrates how Exploratory Data Analysis (EDA) can be effectively used to uncover meaningful insights from poverty-related data across countries and years. By leveraging Python’s powerful data science libraries—Pandas, Matplotlib, and Seaborn—we analyzed key poverty indicators such as the percentage of populations living below international poverty lines (e.g., $2.15, $6.85, and $10 per day).

# Through various visualization techniques, including bar plots, scatter plots, line graphs, heatmaps, box plots, histograms, and correlation heatmaps, we explored the distribution, variability, and trends in global poverty. These analyses provided both a macroscopic view of worldwide poverty patterns and microscopic insights into individual country progress over time.

# Key takeaways from the analysis include:

# Countries in sub-Saharan Africa remain the most affected by extreme poverty.

# Nations like Albania have shown a steady decline in poverty, indicating socioeconomic development.

# Heatmaps and correlation plots revealed how different poverty thresholds are closely interrelated and often aligned with income levels.

# The box and histogram visualizations highlighted the uneven distribution of poverty across countries in specific years.

# This data-driven approach not only enhances our understanding of global poverty but also provides a foundation for informed policymaking, targeted aid allocation, and monitoring of development progress. The project reinforces the importance of combining statistics with visualization to interpret complex global challenges effectively.

# 

# ****6. FUTURE SCOPE****

# While this project provides valuable insights into global poverty trends using exploratory data analysis, there remains significant potential to expand and enhance the analysis in future work. The following areas highlight possible directions for future development:

# 1. Incorporating More Indicators

# Future analyses can include additional socioeconomic indicators such as:

# GDP per capita

# Unemployment rates

# Literacy rates

# Access to clean water, healthcare, and education

# These can help build a more comprehensive picture of the factors influencing poverty.

# 2. Regional and Continental Comparison

# Grouping countries by continent or income level (e.g., low-income, middle-income, high-income) could enable:

# Comparative analysis across regions

# Better understanding of regional challenges

# More focused policy recommendations

# 3. Predictive Modeling

# Machine learning techniques can be introduced to:

# Forecast poverty trends

# Identify potential high-risk countries

# Simulate the impact of interventions (e.g., cash transfers, job programs)

# 4. Time-Series Forecasting

# Use models like ARIMA, Prophet, or LSTM to predict how poverty levels might change in the future under different scenarios.

# 5. Interactive Dashboards

# Develop an interactive dashboard using tools like Plotly Dash, Streamlit, or Power BI to:

# Allow users to explore the data dynamically

# Filter by region, year, or poverty threshold

# View real-time updates and visual comparisons

# 6. Policy Recommendation Engine

# Based on the data, algorithms can be designed to:

# Suggest targeted interventions

# Highlight critical problem areas

# Recommend resource allocation strategies

# 7. Integration with AI for Deeper Insights

# Future systems can use Natural Language Processing (NLP) to analyze related reports and policies, or computer vision to analyze satellite imagery and link it with poverty indicators.

# 

# ****7.REFERENCES****

**Our World in Data – Poverty and Income Inequality**Roser, M., Ortiz-Ospina, E., & Ritchie, H.  
*Available at:* https://ourworldindata.org/poverty  
*Description:* Primary source of the dataset used in this project, providing historical and cross-country data on global poverty indicators.

**World Bank – Poverty & Equity Data Portal**:*Available at:* <https://povertydata.worldbank.org/>  
*Description:* Provides international poverty data including poverty headcount ratios and income thresholds.

**Pandas Documentation**:*Available at:* https://pandas.pydata.org/docs/  
*Description:* Reference for data manipulation techniques used in the project.

**Matplotlib Documentation**  
*Available at:* https://matplotlib.org/stable/contents.html  
*Description:* Used for creating plots such as histograms, line charts, and scatter plots.

**Seaborn Documentation**:*Available at:* https://seaborn.pydata.org/  
*Description:* Used for creating advanced statistical visualizations like heatmaps, box plots, and correlation maps.

**UN Sustainable Development Goals – Goal 1: No Poverty**  
*Available at:* <https://sdgs.un.org/goals/goal1>  
*Description:* Global goal setting and contextual information on poverty reduction initiatives.

# ****Linkdein Link:-**** <https://www.linkedin.com/posts/gowthamgunda_datascience-python-eda-activity-7318266545741185024-gog-?utm_source=share&utm_medium=member_android&rcm=ACoAAEPBLVoBXMNV1wfcw2DZZ9ECmsMBrpy-S7c>

# ****Github Link:-**** <https://github.com/gowthamgunda1045/python-project>













