**Data source:1**

**Data Cleaning & Documentation – FDI by Economic Activity (Panama, 2017–2023)**

**Source**: INEC Panama – Table 7  
**Topic**: Foreign Direct Investment (IED) by Economic Activity  
**Unit**: Thousands of Balboas (B/.)  
**Coverage**: Annual data, 2017 to 2023 (as of December 31 each year)  
**Format**: Excel  
**Primary Key**: Combination of *Economic Sector* and *Year*

**🔍 Raw Data Overview**

The dataset includes FDI values by economic sector along with sector participation and year-over-year variation for 2023.

**Data Cleaning Summary – *FDI by Economic Activity (Panama)***

To prepare the dataset for analysis, I first removed irrelevant rows (1–9, 11, 30–34) that contained notes or were entirely empty. I promoted the correct header row, which included the years 2017–2023, to use as column names.

I renamed the first column to **"Economic Sector"** for clarity, then filtered out rows containing non-data text like explanatory notes and footnotes (e.g., rows 19–21). I also excluded data from 2020 to 2023 that appeared inconsistent or labeled as provisional.

Next, I removed the last two columns—participation percentage and year-over-year variation—since they weren't needed for our core analysis and could be recalculated later if needed.

I checked for and confirmed there were no duplicate rows. Then I used *fill down* to complete any missing sector names, corrected the data types (ensuring years were numeric and FDI values were in number format), and unpivoted the dataset from wide to long format for easier analysis.

Finally, I renamed the columns to:

* **Economic Sector**
* **Year**
* **FDI Economic value**

The cleaned dataset now has **119 rows and 3 columns**.

Cleaned file name: **FDI by Economic Activity (Panama, 2017–2023)**

**Data Source 2:**

**Foreign Direct Investment (FDI) by Country of Origin – Panama (2017–2023)**

* **Source**: INEC Panama – Table 2
* **Topic**: FDI Position by Country of Origin (Top 10)
* **Unit**: Thousands of Balboas (B/.)
* **Period Covered**: 2017 to 2023 (as of December 31)
* **Format**: Excel
* **Primary Key**: Composite — *Country + Year*

**🧹 Data Cleaning Summary**

1. **Removed irrelevant rows and notes**: Rows 1–8, 10, and 24–28 contained metadata, footnotes, or were completely empty — not needed for analysis.
2. **Promoted header row**: Used the row containing year labels (2017–2023) as column headers to enable structured time-series analysis.
3. **Dropped percentage columns**: Columns 9 and 10 (participation %, variation %) were removed — can be recalculated dynamically when required.
4. **Deleted extra empty columns**: Columns 11 and 12 had no data and were dropped for a cleaner scheme.
5. **Removed total FDI rows**: National-level totals were excluded, as we’re focusing on country-level breakdowns and can compute totals separately.
6. **Excluded 'Otros países'**: Limited scope to only the top 10 countries for focused analysis.
7. **Checked for duplicates**: Ensured no duplicate entries for the same country-year combination.
8. **Applied fill-down**: Filled missing country names in merged cells for consistency.
9. **Validated data types**: Ensured country = text, year = integer, value = numeric.
10. **Renamed ‘País de origen’ to ‘Country’**: Aligned with English-language reporting standards.
11. **Corrected spelling inconsistencies**: Standardized country,year names for consistency across the table.
12. **Unpivoted yearly columns**: Transformed the table from wide to long format — one row per country-year entry — to support easy analysis.
13. **Renamed columns for clarity**: Renamed unpivoted fields as *Year* and *FDI\_Value*.

**✅ Cleaned Dataset Overview**

* **Final Columns**: 3 — *Country, Year, FDI\_Value*
* **Total Rows**: 70 (10 countries × 7 years)
* **Cleaned file name**: FDI Position by Country of Origin (Top 10)

**Data Source 3**

**Dataset Summary: Gross Value Added (GVA) by Economic Sector – Panama (2018–2022)**

**Source**: INEC Panama  
**Topic**: Sector-wise Gross Value Added (Valor Agregado Bruto) at basic prices  
**Reference Year**: 2018 (for price adjustments)  
**Format**: Excel (Table 2)  
**Years Covered**: 2018 to 2022

**Data Cleaning Summary: Gross Value Added (GVA) by Economic Sector**

1. **Removed metadata and notes** from column 1 and eliminated the first 6 rows and the last 13 rows from columns 2 to 7 due to nulls and non-data text.
2. **Dropped non-data rows** (e.g. rows 6, 9, 14, and 23) used for category labels or totals to retain only usable records.
3. **Promoted proper headers**, aligning column names with sector names and year values for clarity.
4. **Renamed column 1 sector- category to Sector** for clear identification of economic activities.
5. **Unpivoted year columns** into long format (Sector, Year, GVA\_Value) to make the data analysis ready.
6. **Standardized sector names** and corrected spelling inconsistencies year and across the sector column to ensure uniform labeling across all records.
7. **Validated data types**: confirmed that Year is numeric and GVA\_Value is a float; also reviewed for alias values and corrected them.
8. **Checked for and removed duplicates** based on unique Sector-Year pairs.
9. **Removed subtotal and total rows**, in row 26 to 29 retaining only granular, sector-level data points for accurate analysis.

**✅ Final Dataset Structure:**

* **Columns**: 3 — Sector, Year, GVA\_Value
* **Rows**: 95
* **Cleaned File Name**: *Gross Value Added (GVA) by Economic Sector*
* to Identify Sectoral Growth Drivers and Investment Opportunities (2018–2022)

**Data source 4**

**Data Cleaning Summary: Panama’s GDP by Economic Activity (1996–2022)**

**Source:** INEC Panama – Cuadro 1  
**Topic:** GDP at Purchaser Prices by Economic Activity  
**Unit:** Millions of Balboas (B/.)  
**Period:** 1996–2022

**Cleaning Actions:**

* **Removed Column 1** containing metadata and notes irrelevant to analysis (e.g., titles, footnotes).
* **Dropped non-data rows**: The first 6 rows (columns 2–29) and the last ~24 rows included totals, labels, or nulls not usable in analysis.
* **Promoted headers** to correctly reflect year and sector labels for alignment and clarity.
* **Renamed key column**: Descripción → Sector for better understanding.
* **Filtered nulls** in column 2, row 7 to clean inconsistencies.
* **Unpivoted year columns** into long format (Sector, Year, GDP\_Value) for efficient analysis.
* **Standardized sector descriptions** and corrected inconsistent labels across years columns 1996- 2022.
* **Validated data types**: Ensured Year is numeric, and GDP\_Value is in float format.
* **Removed duplicates and totals** to focus on granular sector-level data. Also applied trimming and check alias rows in numeric columns for accuracy.

**✅ Final Dataset Structure:**

* **Columns**: 3 — Sector, Year, GDP\_Value
* **Rows**: 433
* **File Name**: *Panama’s GDP by Economic Activity (1996–2022)*

**Insights You Can Derive:**

* Fastest-growing sectors post-COVID (e.g. logistics, construction, fintech, tourism).
* Underperforming or shrinking sectors (for risk mitigation).
* Sectors contributing the most to GDP and employment (for policy and investment alignment).

**Data source 5**

**Data Cleaning Summary: *Panama’s Annual GDP (2018–2023)***

**Source:** INEC Panama – PIB Anual  
**Topic:** GDP by Economic Activity (Annual)  
**Unit:** Millions of Balboas (B/.)  
**Period Covered:** 2018–2023  
**Format:** CSV

**✅ Cleaning Procedures**

* **Removed metadata and non-tabular content:**  
  Dropped irrelevant titles, footnotes, and calculated totals from columns 1–2 and the last 3 rows.
* **Eliminated null values:**  
  Removed empty rows (e.g., column 2, row 1) to maintain data integrity.
* **Promoted and renamed headers:**  
  Promoted header rows to reflect actual column names. Renamed "Descripción" to sector for clarity.
* **Standardized text fields:**  
  Trimmed white spaces, corrected spelling inconsistencies, and ensured consistent naming across sectors.
* **Unpivoted the data:**  
  Transformed year columns (2018–2023) into a long-format structure with columns: sector, Year, and GDP\_Value.
* **Validated data types and values:**  
  Ensured Year is integer and GDP\_Value is float. Checked for formatting issues like misplaced commas or decimal points.
* **Filtered totals and duplicates:**  
  Removed aggregated total rows and checked for duplicate sector-year entries.

**📊 Final Dataset**

* **Columns:** 3 → sector, Year, GDP\_Value
* **Rows:** 139 (cleaned and analysis-ready)
* **File Name:** *Panama’s Annual GDP (2018–2023)*

**Data source 6**

**Data Cleaning Summary: *Panama’s Provincial GDP (2018–2023)***

**Source:** INEC Panama – PIB Provincial  
**Topic:** GDP by Province  
**Unit:** Millions of Balboas (B/.)  
**Period Covered:** 2018–2023  
**File Format:** CSV

**✅ Cleaning Steps:**

* **Removed Metadata & Non-Tabular Content:**  
  All introductory rows, footnotes, and subtotals in **column 1** were excluded to isolate the actual dataset.
* **Filtered Sector Rows:**  
  In **column 2**, any rows containing national sectors or categories not related to specific provinces were removed — to focus only on GDP by province.
* **Promoted & Renamed Headers:**  
  The row with actual column names was promoted. The column "Descripción" was renamed to **province** for clarity.
* **Dropped Irrelevant Columns:**  
  Retained only columns with year-based GDP values and the province name; any non-essential columns were excluded.
* **Reshaped the Data (Unpivoted):**  
  Year columns (2018–2023) were transformed into a **long format** with three columns: province, Year, GDP\_Value — enabling easier time-series analysis.
* **Standardized Province Names:**  
  Cleaned and harmonized province names by fixing typos, trimming whitespace, and ensuring uniform formatting across all rows.
* **Validated Data Types:**  
  Confirmed that **Year** is stored as an integer and **GDP\_Value** as a float. Cleaned number formatting issues (e.g., commas, decimal symbols).
* **Removed Duplicates:**  
  Duplicate records were removed, and any total or summary rows were filtered out — keeping only value records , year-by-year entries for each province.

**📊 Final Dataset Overview:**

* **Columns:** 3 — province, Year, P\_GDP\_Value
* **Rows:** 60
* **File Name:** *Panama’s Provincial GDP (2018–2023)*

**Data source 6**

**Data Cleaning Summary: *Panama’s Quarterly GDP (2018–2023)***

**Source:** INEC Panama – PIB Trimestral  
**Topic:** Quarterly GDP by Economic Activity  
**Unit:** Millions of Balboas (B/.)  
**Period Covered:** 2018–2023  
**File Format:** CSV

**✅ Cleaning Steps:**

* **Removed Metadata and Footnotes:**  
  The first few rows (columns 1 and 2) contained metadata like titles, descriptions, and footnotes that were not part of the tabular data. These rows were removed to ensure the dataset begins with clean, structured values suitable for analysis.
* **Promoted Headers & Renamed Columns:**  
  Promoted the row with the actual Year-Quarter values as headers. Renamed "Descripción" to sector to make the column name more intuitive and analysis-friendly.
* **Excluded Non-Granular Totals and Aggregates:**  
  Rows like "Valor Agregado Bruto", "Impuestos sobre la producción netos" and "Producto Interno Bruto a precios de comprador" (rows 19, 20, 21) were removed. These rows are national-level totals and not individual sector activities. Including them would have caused duplication or skewed sector-level analysis.
* **Filtered Only Sector-Level Records:**  
  Removed totals, summaries, or duplicate groupings to ensure the dataset focused solely on specific economic activity categories. This makes time-series and comparative sector analysis more precise.
* **Unpivoted Wide Data to Tidy Format:**  
  Converted quarter-based columns into rows (long format), resulting in three columns: sector, Quarter, and Q\_GDP\_Value. This structure supports easier filtering, visualization, and aggregation.
* **Standardized Sector Names:**  
  Cleaned up inconsistent spelling, extra spaces, and symbols — for example, "Otra producción no de mercado (1)" in column row 18— to improve consistency, avoid duplicate groupings, and simplify data filtering.
* **Validated Data Types:**  
  Ensured that the Quarter field is in text format (e.g., 2018Q1) for consistency and Confirmed that Q\_GDP\_Value is numeric (float), removing thousands of separators and handling currency symbols to ensure smooth calculation and aggregation.
* **Removed Duplicates and Total Rows:**  
  Checked for duplicate records (e.g., repeated sectors for the same quarter) and removed subtotal or total rows to maintain one clean, unique entry per sector per quarter.

**📊 Final Dataset Overview:**

* **Columns:** 3 — sector, Quarter, Q\_GDP\_Value
* **Rows:** 486 ( sector-by-quarter records)
* **File Name:** *Panama’s Quarterly GDP (2018–2023)*