**Agency–Contractor Relationship Strength**

**Build, Clean, Score, and Market-Intel Analysis — Full Walkthrough**

* **What I built:** a clean dataset of agency–vendor relationships + a scoring model + a vendor-focused market analysis pack.
* **Scripts:** Python scripts that (1) assemble/clean columns, (2) compute scores, and (3) create analysis tables & charts.
* **What we get:** Excel workbooks and PNG charts that answer “where is the money?”, “who are the key/competing vendors?”, and “where are opportunities?”

**A. Inputs, Outputs, and Artifacts**

**Source inputs**

* clean\_fpds\_selected.xlsx

A cleaned subset of FPDS-style fields with the exact column names we standardized:

* + **Unique Entity ID**, **Vendor Name**, **Action Date**, **Obligation Amount**, **Awarding Agency**, **City of Performance**, **State of Performance**

This file was built earlier from the source dataset from FPDS. It has one row per **contract action**.

**Scored relationship output**

* agency\_contractor\_scores.xlsx

One row per **Awarding Agency × Vendor** pair with the features and the **Relationship Score (0–100)**. Core columns:

* + **Unique Entity ID** (UEI)
  + **Vendor Name**
  + **Awarding Agency**
  + **Recent Dollars 12m** (sum of obligations in the last 12 months)
  + **Touches 12m** (count of actions in the last 12 months)
  + **Tenure Years** (first action date → last action date, in years)
  + **Recent 18m Flag** (any action in last 18 months)
  + **rank\_dollars / rank\_tenure / rank\_touches** (percentiles **within the same agency**)
  + **Relationship Score (0–100)**

**Market-intel outputs (produced later)**

* vendor\_market\_intel.xlsx

Multi-tab workbook with agency market size, vendor shares, HHI (concentration), vendor tiering, and three “opportunity lists”.

* Charts (PNGs):
  + chart\_top\_agencies\_12m.png
  + chart\_tiers\_by\_agency.png
  + chart\_score\_hist.png
  + chart\_dollars\_vs\_touches.png

These are built only from agency\_contractor\_scores.xlsx (no raw transactions needed at this stage).

**B. What each script does**

**1) Data\_Organizer.py**

**(earlier step; purpose recap)**

* Reads the raw input.
* Selects & renames **only** the required columns to our standard names (listed above).
* Outputs a clean file (e.g., clean\_fpds\_selected.xlsx).

**Why:** ensure consistent column names & types so all following code is simple and robust.

**2) Data\_Analysis.py**

**(Relationship features & scoring)**

**Purpose:** Turn action-level records into **agency–vendor features** and compute the **Relationship Score**.

**Key steps (plain language):**

1. **Read & normalize**
   * Parse dates (Action Date) → true dates (no timezones).
   * Parse money (Obligation Amount) → true numbers (handle $, ,, and ( ) negatives).
   * Build vendor\_key = UEI (fallback to vendor name if UEI missing).
2. **Time windows**
   * Define **today**, and compute two cutoffs:
     + **12 months** for “recent” dollars and touches.
     + **18 months** for the **recency flag**.
3. **Aggregate to Agency × Vendor**
   * **Recent Dollars 12m:** sum of obligations with Action Date in last 12 months.
   * **Touches 12m:** count of actions in last 12 months.
   * **Tenure Years:** (last\_action\_date − first\_action\_date) / 365.25.
   * **Recent 18m Flag:** True if any action in the last 18 months.
4. **Percentile ranks (within agency)**
   * rank\_dollars := percentile of **Recent Dollars 12m** among vendors in the **same agency**.
   * rank\_tenure  := percentile of **Tenure Years** among vendors in the same agency.
   * rank\_touches := percentile of **Touches 12m** among vendors in the same agency.

Rules:

* + If **only one value** exists → rank = 0.00 (avoid false “#1”).
  + If **all values identical** → rank = 0.50 (neutral).
  + Otherwise use percentile (rank(pct=True, method="average")).

1. **Relationship Score (0–100)**
   * Combine ranks with weights:

raw = 0.50\*rank\_dollars + 0.30\*rank\_tenure + 0.20\*rank\_touches

* + Penalty for inactivity:

if Recent 18m Flag == False: raw \* 0.85

* + Scale to 0–100 and round to 2 decimals.

1. **Save** → agency\_contractor\_scores.xlsx (+ CSV)

**Why these weights?** (business logic)

* **Dollars (50%)** signal strategic importance.
* **Tenure (30%)** signals trust & history.
* **Touches (20%)** signals operational engagement.
* **Penalty** prevents dormant relationships from ranking as “strong” purely on history.

**3) Data\_Analysis\_Main.py**

(uses *only* agency\_contractor\_scores.xlsx)

**Purpose:** Produce a **vendor-facing market intel pack** for opportunity scouting.

**What it computes:**

1. **Agency Market Size (12m)**
   * Sum Recent Dollars 12m per agency.
   * Compute **Agency Share of Total** = Agency 12m / Sum(all agencies 12m).
   * Output tab + chart chart\_top\_agencies\_12m.png.
2. **Competitive Landscape (12m)**
   * Per agency, sum **Vendor Obligations 12m** (i.e., vendor’s 12m dollars).
   * Compute **Vendor Market Share 12m** = Vendor 12m / Agency 12m.
   * Make **Top 10 Vendors per Agency**.
   * Compute **Herfindahl–Hirschman Index** (HHI - concentration) = Σ(share²).
     + Read as:
       - **<1500** unconcentrated (open/competitive),
       - **1500–2500** moderate,
       - **>2500** concentrated (few dominate).
3. **Vendor Segmentation (tiers)**
   * Tier logic (per Agency × Vendor):
     + **Elite**: score ≥ 80 and **recent**
     + **Core**: 60–80 and **recent**
     + **Emerging**: 40–60 or (score<60 and **Touches ≥ 2**)
     + **Prospect**: score < 40 and **recent**
     + **Dormant**: **not** recent
   * Output tab: “Vendor Tiers”
   * Chart: chart\_tiers\_by\_agency.png (stacked counts by agency)
4. **Opportunity Lists (actionable)**
   * **Re-engage Targets:** Dormant **and** (Tenure ≥ 1y **or** rank\_dollars in top quartile).
   * **Emerging Targets:** Tier=Emerging **and** (Touches ≥ 2 **or** rank\_touches ≥ 0.6).
   * **Greenfield Targets:** Agencies with low HHI; pick vendors with **low score (<40)** and **tiny share (<5%)**.
   * Output tabs for each.
5. **Score Distributions & Drivers**
   * **Histogram** of Relationship Score → chart\_score\_hist.png.
   * **Scatter** Recent Dollars 12m vs Touches 12m (bubble=Tenure) → chart\_dollars\_vs\_touches.png.

All the above are written into vendor\_market\_intel.xlsx, plus the PNGs.

***C. How to run (Mac / Terminal) –* For Technical Purpose**

*From your project folder:*

*# 1) Relationship features + score (reads clean\_fpds\_selected.xlsx, writes agency\_contractor\_scores.xlsx)*

*python3 "Data Analysis/Data\_Analysis.py"*

*# 2) Market intel pack (reads agency\_contractor\_scores.xlsx, writes vendor\_market\_intel.xlsx + PNGs)*

*python3 "Data Analysis/vendor\_market\_intel\_from\_scores.py"*

*Important: paths with spaces need quotes, e.g., "Data Analysis/...".*

**D. Data dictionary (columns you’ll see and how to read them)**

| **Column** | **Meaning** | **How we compute it** |
| --- | --- | --- |
| **Awarding Agency** | Agency (buyer) | From input |
| **Unique Entity ID** | UEI | From input |
| **Vendor Name** | Vendor’s legal name | From input |
| **vendor\_key** | Stable vendor key | UEI; fallback: uppercase name |
| **Action Date** | Date of action | From input |
| **Obligation Amount** | $ amount for action | From input |
| **Recent Dollars 12m** | Sum of obligations in last 12 months | Filter by Action Date |
| **Touches 12m** | Count of actions in last 12 months | Filter by Action Date |
| **Tenure Years** | Relationship length in years | (max\_date − min\_date)/365.25 |
| **Recent 18m Flag** | Has any action in last 18 months? | Boolean |
| **rank\_dollars** | Percentile of 12m dollars **within same agency** | rank(pct=True) w/ tie & edge rules |
| **rank\_tenure** | Percentile of tenure **within agency** | Same |
| **rank\_touches** | Percentile of touches **within agency** | Same |
| **Relationship Score (0–100)** | Composite strength score | Weighted ranks + inactivity penalty |

**E. Formulas (for reference)**

**Vendor Market Share (12m, inside an agency)**

Share = (Vendor Obligations 12m) / (Agency Total Obligations 12m)

**Agency Share of Total**

Agency Share = (Agency Total 12m) / ( Sum of Agency Total 12m)

**HHI (Herfindahl-Hirschman Index)**

HHI = 10,000 \* (Sum of Square of Vendor Shares)

**Relationship Score**

Raw = (0.50 \* rank\_dollars) + (0.30 \* rank\_tenures) + (0.20 \* rank\_touches)

**F. How to read the charts you attached**

**1) Top Agencies by Spend (12m)**

**chart\_top\_agencies\_12m.png**

* **What it shows:** which agencies obligated the most in the last 12 months (from our dataset).
* **Use it for:** picking target agencies.

**2) Vendor Tiers by Agency**

**chart\_tiers\_by\_agency.png**

* **What it shows:** how many **Elite/Core/Emerging/Prospect/Dormant** vendors each agency has.
* **Use it for:** seeing ecosystem health; choose agencies with more **Emerging/Prospect** (open), or audit agencies with many **Dormant** (re-engage list).

**3) Score Distribution (All Agencies)**

**chart\_score\_hist.png**

* **What it shows:** how relationship scores spread overall.
* **Use it for:** identify mid-tier clusters; set thresholds for sales focus.

**4) Recent Dollars vs Touches (bubble = Tenure)**

**chart\_dollars\_vs\_touches.png**

* **What it shows:** how money and interaction volume relate, with bubble size showing relationship length.
* **Use it for:** understand what drives score in your dataset; spot “big-award few-touch” vendors vs steady operators.

**G. Files you can hand off**

* **Input:**

clean\_fpds\_selected.xlsx (action-level, standardized columns)

* **Relationship score output:**

agency\_contractor\_scores.xlsx (one row per Agency × Vendor)

* **Market-intel pack:**

vendor\_market\_intel.xlsx with tabs:

* + Read Me
  + Agency Market 12m
  + Vendor Share 12m
  + Top10 Vendors per Agency
  + Market Concentration (HHI)
  + Vendor Tiers
  + Re-engage Targets
  + Emerging Targets
  + Greenfield Targets
* **Charts:**

chart\_top\_agencies\_12m.png, chart\_tiers\_by\_agency.png,

chart\_score\_hist.png, chart\_dollars\_vs\_touches.png