

Cash-Pay Healthcare Market Analytics

This repository provides a multi-angle analysis of the U.S. **cash-pay healthcare market** based on physician-level payment data.

Through robust data preprocessing and three complementary analytical plans (A, B, C), we reveal actionable insights for startups like **Healium** to guide expansion, pricing, and marketing decisions.









Project Background

As the U.S. healthcare landscape evolves, a growing number of patients are seeking **cash-pay services** — bypassing traditional insurance in favor of direct, transparent care.

This project aims to identify **where and how** these markets are emerging using open data from [CMS Open Payments](#).



Files in This Repo

Filename	Description
cashpay_data_preprocessing_core.ipynb	 Cleans raw data from <code>data.csv</code> , filters 2023 & cash-based records, generates <code>state_agg.csv</code> for all plans
PlanA_steady_market_evaluation.ipynb	 Calculates log-scaled hotness scores to identify structurally stable cash-pay states
PlanB_high_value_clinic_targets.ipynb	 Uses average payment per transaction to target premium opportunity markets
PlanC_state_market_clusters_KMeans.ipynb	 Uses KMeans clustering to group states into three market types (broad, boutique, volatile)
data.csv	 Raw data file containing 2023 CMS Open Payments transaction-level records
state_agg.csv	 Aggregated state-level dataset generated from <code>cashpay_data_preprocessing_core.ipynb</code>






Plan Overview

Plan	Core Metric(s)	Focus	Suitable For
A	<code>Num_Records</code> , <code>Total_Payment</code> (log)	Stable activity, reduced outliers	Steady growth & budget-safe expansion

Plan	Core Metric(s)	Focus	Suitable For
B	Avg_Payment	Value per interaction	Premium product targeting, concierge care
C	All 3 (KMeans Clustering)	Market segmentation	Strategic planning by market type

Example Visual Outputs

-  Top 10 States by Hotness Score (Plan A)
-  Average Payment Distribution (Plan B)
-  U.S. Market Type Heatmap (Plan C)

All visuals are generated using Matplotlib, Plotly, and Seaborn inside each notebook.

How to Use

1. Clone or download this repo
2. Start with `cashpay_data_preprocessing_core.ipynb` to prepare the dataset
3. Run any of the analytical notebooks (A/B/C) independently based on your focus
4. Use exported CSVs or visual charts in presentations, internal planning, or market entry strategy

Data Source

- [CMS Open Payments](#)
- Filtered for:
 - Year: **2023**
 - Payment Type: **Cash or Cash Equivalent**
 - Fields used: Physician State, Payment Amount, etc.

Special Notes

- This project explores the cash-pay healthcare market, a rapidly growing segment in U.S. healthtech. It's particularly relevant to startups like **Healium** that aim to optimize provider discovery, price transparency, and market targeting in self-pay contexts.
- All clustering and scoring logic is **interpretable** and aligned with business applications.
- Future improvements can include: time-series trends, provider-level drill-down, or provider specialty segmentation.

Author

Stefan Su

Tech-Business Hybrid | Data Analyst & Builder | [LinkedIn](#)

 Contact for collaboration, strategy consulting, or startup ideas.