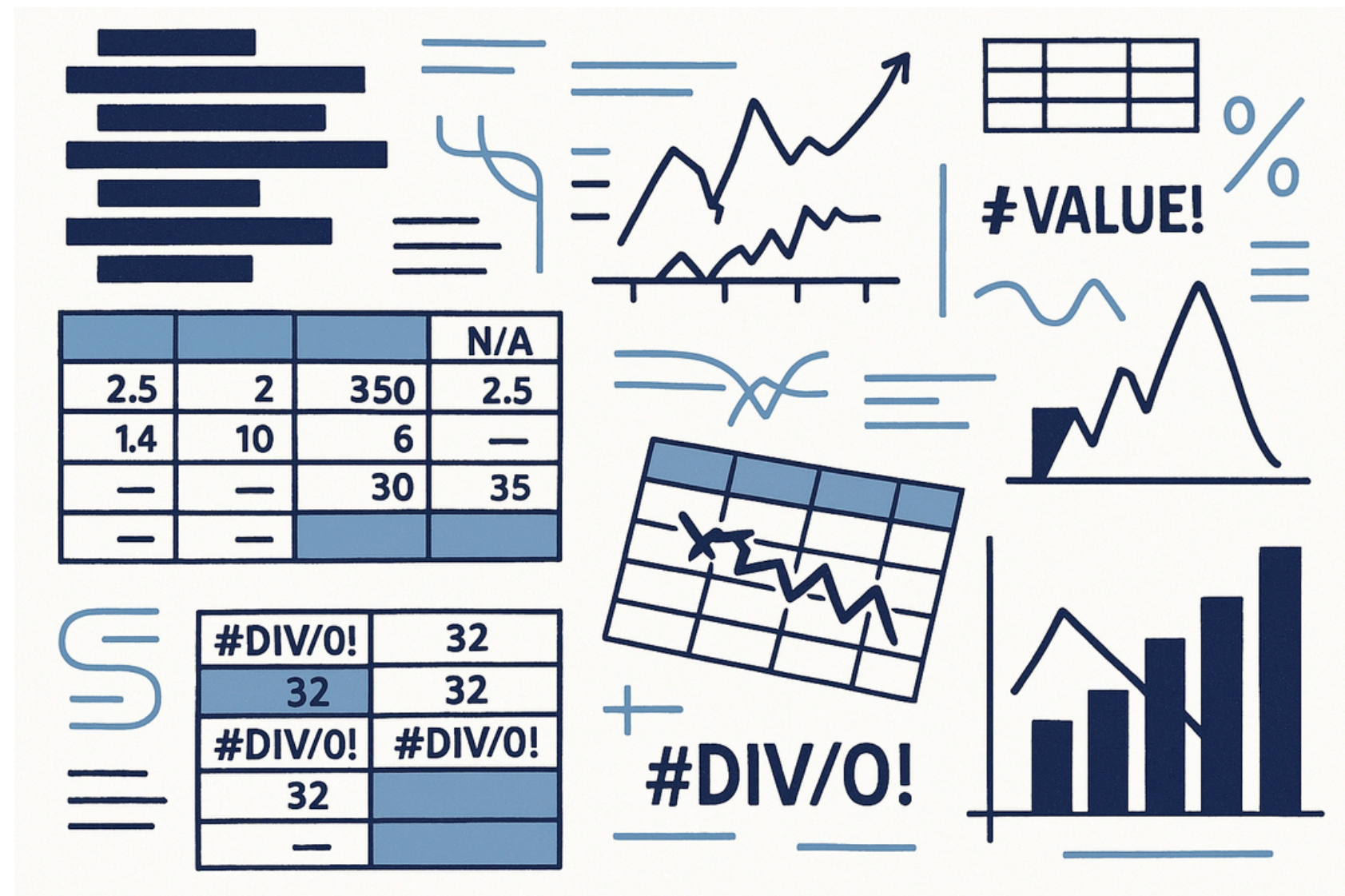# Arch's Mission.
My Contribution.

Harsh Sharma | Github | harshnarayan.sharma@duke.edu

The Risk of Good Enough Data

- The world of private investing is full of data but it is often unusable.
- Inconsistent formats, missing details, and manual errors prevent a clear view of performance.
- It creates a liability for clients and a bottleneck for the business.

A Closer Look at the Risk

FILINGMANAGER_ZIPCO...	REPORTTYPE	FORM13FFILENUMBER	CRDNUMBER	VALUE
04538-133	13F NOTICE	028-15342	159875	null
77057	13F HOLDINGS REPORT	028-23667	null	287125
77057	13F HOLDINGS REPORT	028-23667	null	231269
77057	13F HOLDINGS REPORT	028-23667	null	275963
77057	13F HOLDINGS REPORT	028-23667	null	210739
77057	13F HOLDINGS REPORT	028-23667	null	549791
77057	13F HOLDINGS REPORT	028-23667	null	1261607
77057	13F HOLDINGS REPORT	028-23667	null	305549
77057	13F HOLDINGS REPORT	028-23667	null	15371607
77057	13F HOLDINGS REPORT	028-23667	null	5637224
77057	13F HOLDINGS REPORT	028-23667	null	2668798
77057	13F HOLDINGS REPORT	028-23667	null	377126
77057	13F HOLDINGS REPORT	028-23667	null	6781310
77057	13F HOLDINGS REPORT	028-23667	null	6670258

- Inconsistent Values: The VALUE column is a mix of dollars and thousands, making financial analysis impossible.
- Data Gaps: Key columns are riddled with missing values, leaving crucial information up to chance.
- Unstructured Chaos: Related data is scattered across multiple files, forcing a manual, time-consuming join.

The Automated Path to Data Integrity

RAW DATA → CLEANING → STANDARDIZING → VALIDATING → CLEAN DATA

My role is to be the Data Quality Associate who builds a foundation of trust. I developed an automated pipeline to transform this raw data into a reliable, single source of truth.

My Solution: The Python Pipeline

Building on a Solid Foundation

ACCESSION_NUMBER	FILING_DATE	SUBMISSIONTYPE	VALUE	CIK
0001569785-25-000006	7/31/2025	13F-NT	0.00	1569785
0001963669-25-000003	7/31/2025	13F-HR	287,125.00	1963669
0001963669-25-000003	7/31/2025	13F-HR	231,269.00	1963669
0001963669-25-000003	7/31/2025	13F-HR	275,963.00	1963669
0001963669-25-000003	7/31/2025	13F-HR	210,739.00	1963669
0001963669-25-000003	7/31/2025	13F-HR	549,791.00	1963669
0001963669-25-000003	7/31/2025	13F-HR	1,261,607.00	1963669
0001963669-25-000003	7/31/2025	13F-HR	305,549.00	1963669
0001963669-25-000003	7/31/2025	13F-HR	15,371,607.00	1963669
0001963669-25-000003	7/31/2025	13F-HR	5,637,224.00	1963669
0001963669-25-000003	7/31/2025	13F-HR	2,668,798.00	1963669
0001963669-25-000003	7/31/2025	13F-HR	377,126.00	1963669
0001963669-25-000003	7/31/2025	13F-HR	6,781,310.00	1963669

- Financial Accuracy: All VALUE metrics are normalized, enabling confident, reliable analysis.
- Complete Data: Critical gaps are filled, providing a complete picture of holdings and reports.
- Actionable Insight: The data is unified into a single, clean file, allowing teams to focus on strategy, not manual work.

Why I'd Love to be Part of the Mission

I absolutely love the idea behind creating this platform of becoming the one stop destination in the alternative investments landscape by through technology and collaboration.

My project was a small taste of my passion for the kind of work that makes that mission possible.

Thank you! I would love the opportunity to contribute to the future you're building.

THANK YOU

Harsh Sharma | Github | harshnarayan.sharma@duke.edu