



A Data Pipeline to Prove Fraud

Based on a true story...



Overview

- Step 1. Exploratory Data Analysis (EDA)
 - 1a. Relevant Domain Knowledge:
 - What context is motivating this analysis
 - 1b. What data might be relevant?
 - 1c. What does initial look at data suggest?
- Step 2. What does the EDA suggest deserves a closer look?



Overview

- Step 3. What does the deeper analysis suggest?
- Step 4. What would a third-party need to do in order to reproduce this analysis?
 - Why is reproducibility important in this context?
 - Hint: What do you think would happen if these analyses were NOT reproducible?

Step 1a. Relevant Domain Knowledge

What context is motivating analysis?

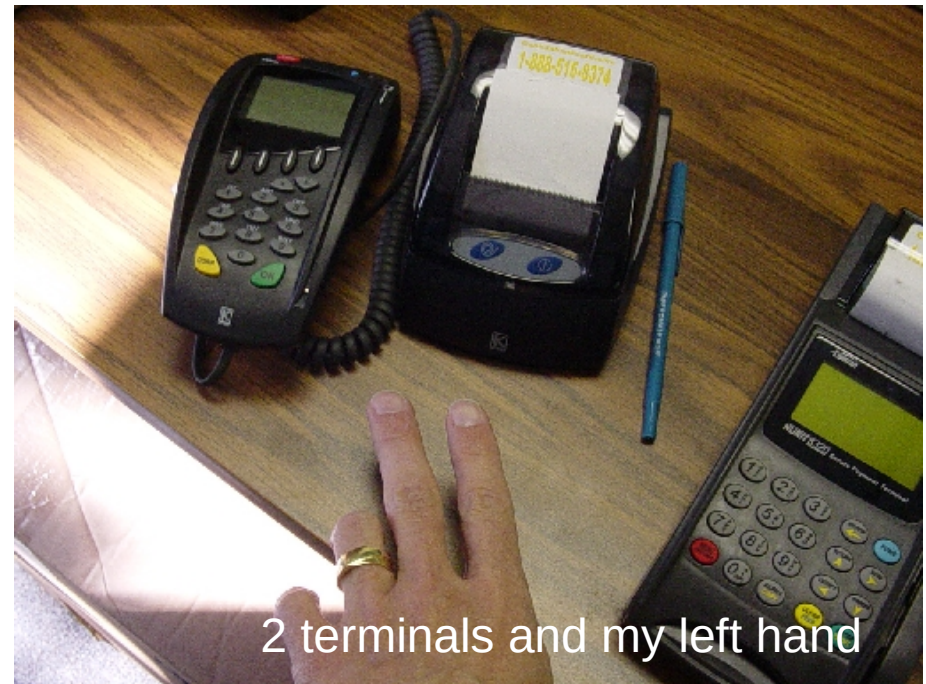
- Investment motivation:
 - Before 2001, only Canadian banks were allowed to provide Point of Sale (POS) credit/debit card services
 - Deregulation opened lucrative opp's for 3rd party providers
 - CBS 2001 Inc. healthy revenue growth, but capital limited
- After 2 months of DD, bought 29 terminals for \$46k in Jan 2004.
 - Term was 5 years. Annualized return was 20%
- CVP founded June 2004 to raise additional capital for CBS 2001
 - 1st client signed on Oct 2004
 - 5 more sign on between April 2005 and July 2006
- July 2006 payment bounces. Aug – Dec: lots of unkept promises

Step 1a. (Cont.) Context motivating analysis

- Demand letter Feb 2007. First action filed March 2007
- Judgment June 2007
 - JDE December 2007: First opportunity to request data
 - What
- CVP founded June 2004
 - 1st client signed on Oct 2004
 - 5 more sign on April 2005 thru July 2006
- July 2006 payment bounces. Aug – Dec: lots of unkept promises

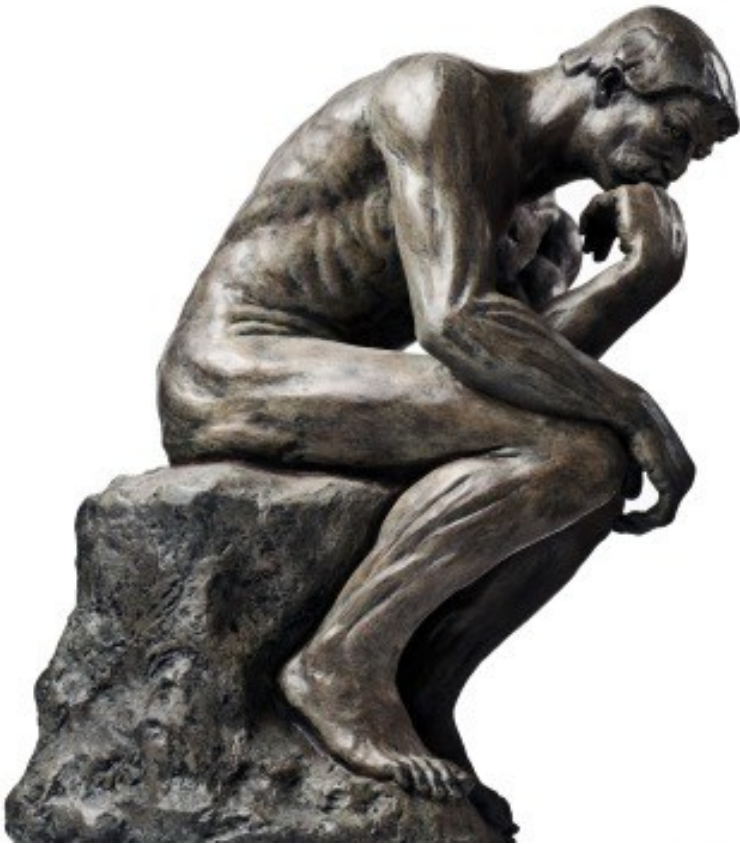
Step 1a. Domain Knowledge (cont.)

- Demand letter Feb 2007. First action filed March 2007
- Judgment June 2007
- What exactly were the investment assets again?



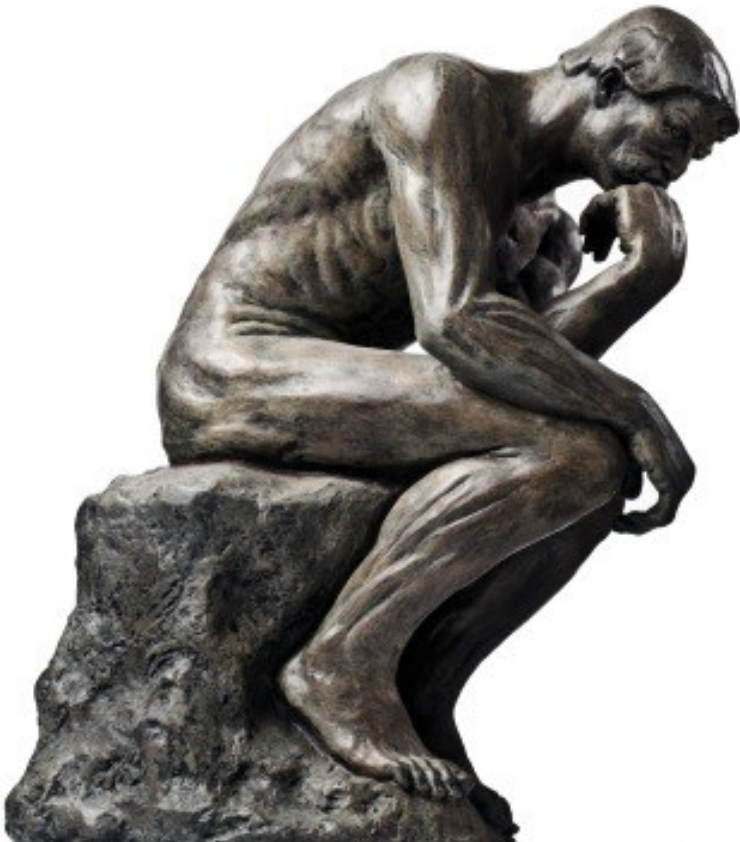
Step 1a. Domain Knowledge (cont.)

- It's June 2007 and you have a judgment. Now what?

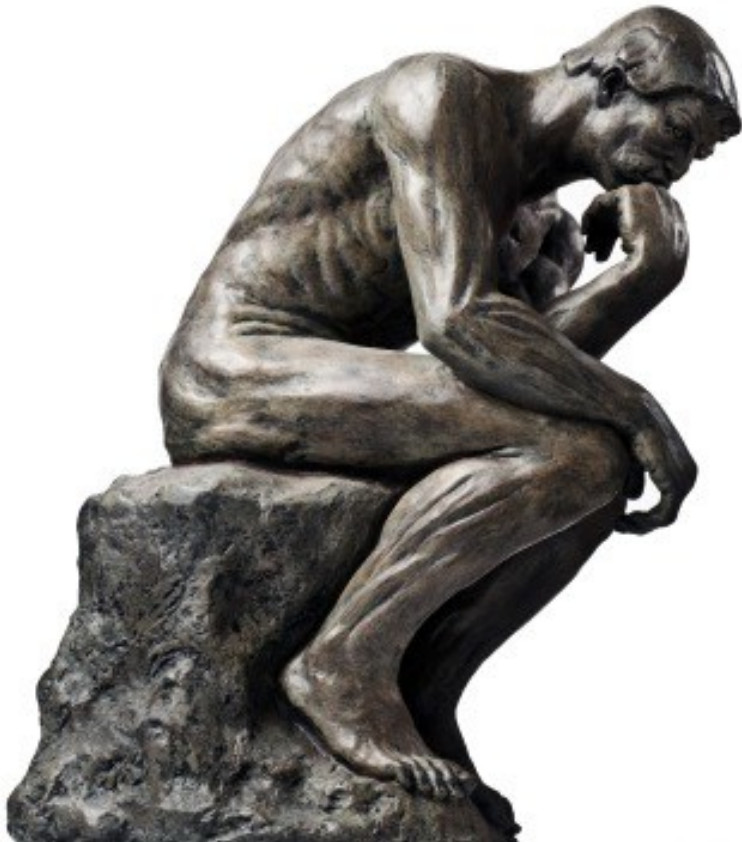


Step 1a. Domain Knowledge (cont.)

- It's June 2007 and you have a judgment. Now what?
- Right! Try to collect on it.

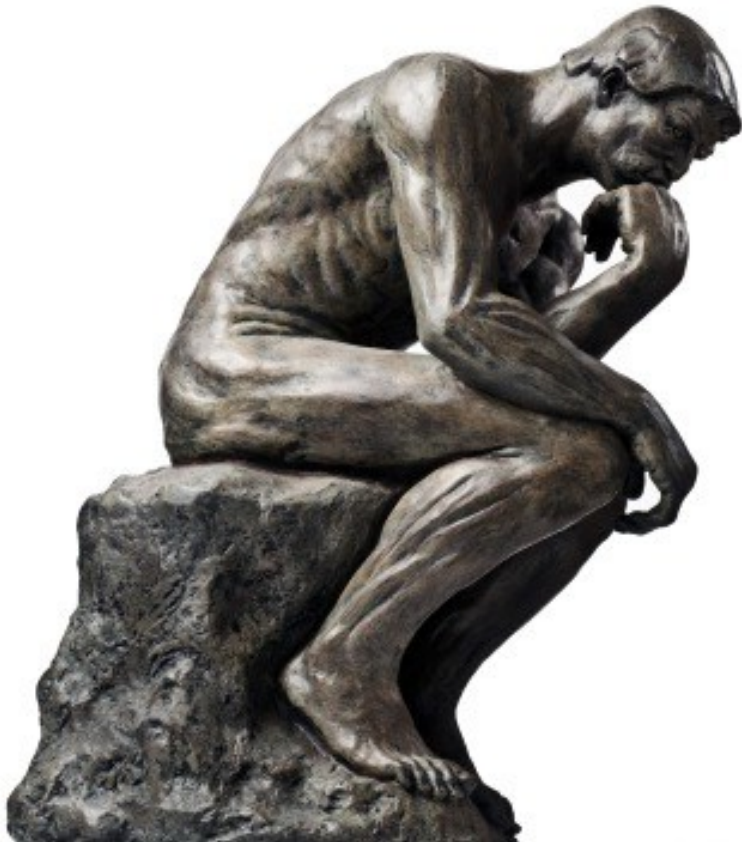


Step 1a. Domain Knowledge (cont.)



- It's June 2007 and you have a judgment. Now what?
- Right! Try to collect on it.
- But what do you think happened when you tried to collect on it?

Step 1a. Domain Knowledge (cont.)



- It's June 2007 and you have a judgment. Now what?
- Right! Try to collect on it.
- But what do you think happened when you tried to collect on it?
- Well, you manage to collect a percent or two using garnishments, but that money stops after a few months.

Step 1b. What data might be relevant?

- You schedule your JDE for Decemeber 2007
 - This will be your 1st chance to request data.
 - What do you ask for?



Step 1b. What data might be relevant?

- You schedule your JDE for Decemeber 2007
 - This will be your 1st chance to request data.
 - What do you ask for?
- Right, you ask for information regarding the assets



Step 1b. What data might be relevant?

- You schedule your JDE for Decemeber 2007
 - This will be your 1st chance to request data.
 - What do you ask for?
- Right, you ask for information regarding the assets
 - Defendant claims company has no money or assets



Step 1b. What data might be relevant?

- You schedule your JDE for Decemeber 2007
 - This will be your 1st chance to request data.
 - What do you ask for?
- Right, you ask for information regarding the assets
 - Defendant claims company has no money or assets
 - Now what?



Step 1b. What data might be relevant?



- Right again! You ask who the company used as their payment processor.
- Request data on merchants the company signed on over relevant time period
 - 1/2004 thru 4/2008 (most recent at that time)
- What did this data look like?
 - See jupyter notebook...