**Position:** Research Data Analyst I/II (JC-359684) Practical Exercise

Confidential – Do not share or distribute

Thank you for applying to work at CDSS! This stage of the application process is an opportunity for you to get a feel for the questions you may be asked to investigate as a researcher in the Research Automation and Data Division.

We encourage you to work with whatever statistical software or analytical tools you are most comfortable with. The goal of the exercise is not to test proficiency in a specific software, but to see how you work with data to answer questions that are important for CDSS programs.

**Plan to spend 3-4 hours working on this task** – **limiting yourself to no more than 4 hours**.This is intendedto help you focus on sharing your thinking rather than writing a lot of code.

While you are expected to follow the prompts below there is not a ‘correct’ answer. In the spirit of exploratory research, you are encouraged to analyze the problem, interpret results, and highlight findings as you see fit.

**Analytical Task: Estimate the change in administrative theft claims and reimbursement payments**

***Background***

CDSS assists millions of low-income Californians through the Electronic Benefits Transfer (EBT) system. EBT cards a like a state-issued debit card that people can use to redeem their public assistance benefits like CalFresh, CalWORKs and other food and cash aid benefits. Like privately issued credit cards, criminal organizations use a range of strategies to compromise EBT card information and use that information to steal people’s benefits. When this happens, CDSS program participants can report their benefits stolen and be reimbursed after a brief investigation.

Until recently, program participants had to take three steps to be reimbursed for EBT theft:

1. File a police report in their local police jurisdiction
2. File an administrative claim with the card issuing company (henceforth, “FIS”) that CDSS contracts to operate the EBT payment system to confirm that there was no payment processing error that resulted in lost benefits
3. File a theft report with their county case worker who confirms the details of the theft transaction and coordinates the investigation into the theft claim.

On January 27, 2023, CDSS released [a policy directive](https://www.cdss.ca.gov/Portals/9/Additional-Resources/Letters-and-Notices/ACLs/2023/23-13.pdf?ver=2023-01-27-165817-090) changing this reporting protocol so that **steps 1 and 2 above are no longer required** with the goal of reducing the administrative burden of theft reporting. CDSS does not have data on police reports, but it does have access to all claims filed with FIS (step 2 above) as well as the number and value of reimbursements following step 3. This policy change went into effect on January 27th across all counties in California.

CDSS Directorate is interested in understanding how the policy change on January 27th affected the volume of theft reimbursement activity.

***Deliverable Instructions***

Write a **two-page memo** to Research, Automation, and Data Division Deputy Director summarizing the findings of your theft reporting and reimbursement analysis.   
(Single space, 1-inch margins, 11 point font)

In your memo, you should address the following:

1. Summary assessment of whether/how administrative claims to the card issuing company and theft reimbursement payments changed after the policy change
2. Model or empirical strategy you used determine the extent of the change
3. Explanation of why you choose the model or empirical strategy
4. Limitations of the data or the empirical strategy chosen

**Upload your two-page memo, code, and final data sets to the “Deliverable Submission” subfolder of the OneDrive folder linked in the email containing these instructions.**

***Data Sources***

You have access to two de-identified administrative data files from CDSS – claims to FIS and theft reimbursement transactions.

*claims.csv*

This file contains individual-level claims filed with the card issuing company FIS between June 1st, 2022 and March 31st, 2023. The file contains the following data elements:

* **CaseID**: encrypted benefit case number, a unique identifier assigned by the state to all assisted individuals in a household.
  + Across this time range, the same individuals would have the same case number, even if their card numbers change.
* **CardID**: encrypted EBT card number, each adult EBT cardholder will have a unique EBT card number. There can be multiple EBT cardholders on a case but EBT card values are unique across cases.
  + The same individual might have different card numbers over time if they lose their card or have their card replaced after being the victim of EBT theft.
* **Case\_county**: county of residence of all individuals associated with a given case
* **Client\_language**: primary spoken language of the individual reporting the theft incident. Different individuals within a case might have different primary spoken languages.
* **Retail\_zip**: zip code of the retailer location where the theft transaction occurred
* **Adj\_amt:** dollar value of the theft transaction being reported
* **Transact\_date**: date on which the theft transaction occurred (not the date the claim was filed with FIS)

*reimbursements.csv*

This file contains individual-level reimbursement payments to EBT card holders by CDSS between June 1st, 2022 and March 31st, 2023. The file contains the following data elements:

* **CaseID**: same as in claims.csv. You can
* **CardID\_X (1-4)**: 4 variables representing the encrypted EBT card numbers for up to 4 individuals on a given caseID.
* **Case\_county**: county of residence of all individuals associated with a given case
* **Client\_language\_X (1-4)**: 4 variables representing the primary spoken language of the individuals on a given caseID. client\_language\_1 corresponds to the spoken language of the cardholder of cardID\_1
* **Benefit\_amount**: dollar value of the reimbursement payment to the victim of EBT theft
  + Note: this does not correspond one-to-one with values of transactions reported in the claims.csv file since multiple theft transactions reported in the claims data might be reimbursed as a single payment in the reimbursements data.
* **Benefit\_year**: year that the stolen benefits being reimbursed were originally issued (loaded onto the victim’s card)
* **Benefit\_month**: month that the stolen benefits being reimbursed were originally issued (loaded onto the victim’s card)
  + Ex: CDSS issues cash benefits to individuals on a CalWORKs case at midnight on 2/1/23. A criminal organization goes to an ATM at 1:00 am on 2/1/23 and withdraws the cash before people on that case can use the benefits. Regardless of when the theft is reported to CDSS, when the reimbursement is processed, the benefit month will be Feburary of 2023, since the stolen benefits were issued in February 2023.
* **Pay\_date**: date that the reimbursement payment to the theft victim was processed (I.e., date they were made whole for the money stolen).

*Notes on how these files relate to one another*

It is not essential for the purpose of this exercise to attempt to link records across files on caseID or cardID Prior the policy change, filing a claim with FIS was a required step before people could be reimbursed for theft, so we would expect a closer relationship between the volume of claim activity and reimbursement activity in the pre-policy period. Even in this time period, however, we wouldn’t expect 100% compliance with the old policy – it's possible that a small number of people may not have filed a claim with FIS but still eventually received a theft reimbursement.

Additionally, not all claims to FIS would have a corresponding theft reimbursement. We would expect significantly more FIS claims for at least three reasons: 1) a single reimbursement payment might cover multiple claims to FIS (as discussed above), 2) some clients may fail to go from step 2 to step 3 in the pre-policy theft reporting protocol for a variety of reasons, and 3) even clients who file a claim then file a report with their county case worker may fail to get their reimbursement request approved. Finally, the reimbursements data do not contain the date, time, and location of the specific transaction made by the criminal organization that stole the benefits, *so there is no precise way to match claims to reimbursements*. You can only see which cases made a claim and received a reimbursement and when.

Your analysis should focus on the extent to which the volume and value of claims and theft reimbursements change following the policy change on January 27th – looking at the connections between cases/individuals in both data sets is not required but could complement your findings.