**2022 Travelers Business Insights & Analytics LDP Case Competition**

**Business Problem:**

You are a recent college graduate that has been scouted by the new insurtech startup company, Trusurance. They believe your experience analyzing datasets and your critical thinking abilities are exactly what they need to stand out from the rest of the insurtechs in the market. Trusurance specializes in property casualty insurance and advertises the ability to save money on claim handling with their cutting-edge fraud detection capabilities.

It is a Tuesday morning, a week after your started. The CEO of the company walks over to your desk and requests your help. Apparently, their cutting-edge fraud detection has not been developed yet and they were hoping you would be the one to develop it. At your disposal is a list of claims and access to standard industry technology. The CEO has already setup a meeting with the CEOs of several major insurance companies and has named you as the key speaker. With your job on the line, you quickly reach out to the rest of your team and devise the next steps.

With your team, analyze the provided dataset. You can use whatever tools you have at your disposal. Additional trainings and reference material will be provided throughout the length of the competition. Based on your research create a method to identify fraudulent auto claims. This could be a simple business rule engine (a series of conditional statements) or as complex as a deep learning artificial intelligence model. Evaluate your fraud detection method by testing it against the provided data. **There is no one solution.** You will have to create a presentation that summarizes your approach and results to a panel of judges. The judges have a variety of different backgrounds so take that into consideration when creating the presentation. These judges represent the CEOs of other companies so you will have to convince them if you want their buy in.

Remember your audience!

**Data Description:**

This dataset contains two files. The first file contains 1572 unique claims. Claims typically have more than one claimant associated to them. The second file contains 1572 associated incidents. Each incident represents an automobile crash of an individual on the claim. They should be assumed to be the policy holder for the purposes of this case.

**Variable Descriptions:**

**Claim\_final.csv:**

|  |  |
| --- | --- |
| CLM\_NBR | A system-generated number that identifies a claim. It is unique within WC and APL. |
| NBR\_OF\_CLMT | A sum of the number of claimants on a claim. |
| BI\_CLMT\_CNMT | A count of the number of claimants on a claim that are for Bodily Injury. |
| COV\_RGST\_DT | The date on which the claim was registered to the policy. |
| CLAIM\_EXPENSE\_EST\_AMT | The amount of the Allocated Loss Adjustment Expense (ALAE) incurred amount summarized at the claim level. |
| POL\_NBR | A number that identifies an insurance policy. |
| POLICY\_STATE | A code that identifies the primary state of an insurance policy |
| PRI\_BTH\_DT | The date of birth of the individual. |
|  |  |

**Claimant\_Final.csv:**

|  |  |
| --- | --- |
| CLM\_NBR | A system-generated number that identifies a claim. It is unique within WC and APL. |
| AUTO\_ACCIDENT\_DESC | This is the accident type as captured at the time of notice of loss by the claim handler. |
| CLAIMANT\_STATUS | A system-generated code that identifies the current processing status of a claimant segment. |
| APPR\_DAM\_EST\_REC\_CNT | The count of damage estimates performed on the claimant's vehicle. |
| VEH\_COLR\_TXT | A description of the color of the claimant's vehicle. |
| VEH\_DAM\_TXT | The text describing the damage description field on the Appraisal request screen. This does not prefill with information from the claimant screen. |
| CLMT\_VEH\_MDL\_NM | A description of the field that displays the model of the vehicle. This a free-form text field. |
| CLMT\_VEH\_YR | The year of the vehicle. |
| VEH\_TYPE | A code that identifies the class of vehicle (i.e. auto or heavy equipment). |
| VEH\_PRIMARY\_PT\_OF\_DAMAGE | A code that identifies the primary point of damage. |
| AIRBAG\_DEPLOYED | A code that identifies how many airbags were deployed during the accident within the vehicle associated to the claimant. |
| VEH\_SPEED\_AT\_IMPACT | A code that identifies how fast the vehicle was traveling at the time of the accident. |
|  |  |

**Analysis & Visualizations:**

There are many different approaches and technologies available to analyze the provided dataset. You may use anything at your disposal if you can present the material in a manner understood by someone who might not be familiar with it. We have identified several pieces of industry standard software to aid you through this endeavor. Regardless of approach, you should be able to derive patterns and correlations which will ultimately be the foundations of your solution. In other words, a simple 2-dimensional chart can be just as powerful as an artificial intelligence model. Therefore, collaborate with your group and understand what skills and tools you are most familiar with that best aid your research and analysis. The key to analytics is understanding your data!

After you conclude your analysis and devise a fraud detection method, you will need to structure your approach, findings, and solution into a presentation. You may use figures, tables, and text throughout the presentation to aid in the judges’ understanding. Remember, the goal of the presentation is to convince the judges of the accuracy of your work. The analysis is the foundation of your conviction but is only part of the overall solution. You will need to translate your solution into business value; often stated as “Why should I care?” which will require additional metrics based on the provided dataset.

Each group may have at **most 3** people and will:

* Work together within the group on data analysis, but not between groups
* Explore the characteristics of the incidents in the dataset and identify fraudulent auto claims
  + You can use any software for this task, but the recommended tool is/are:
    - SQL
    - Python (Jupyter Notebook, VSCode, in any text editor of your choice)
    - Excel
* Build your story of the data using a dashboard
  + You can use any software for this task, but the recommended tool is/are:
    - Excel
    - Tableau
    - QlikView
    - MicroStrategy
  + No limitation on the types of visualizations/charts/tables
* Prepare presentation slides to summarize the analysis, your fraud detection method, and the business value of your solution.

**Data Documentation:**

The following is a guideline on what to include in the documentation:

1. Prepare a document with the following:

* All fields in the final table, whether they were initially provided or created
* Logic applied to obtain the field or directly taken from source provided
* Fields used in creations of new field
* Description of field
* Indicator to determine if field is part of key that makes each record unique
* Data types
* Comments or Observations

2. Analysis & Visualization explanation. These can include, but are not limited to:

* What do the different Graphs, Metrics, Visualizations, etc., mean?
* How are the different items in your Analysis & Visualizations created?

**Timeline:**

Case Competition Kick Off: Wednesday, February 15th

Registrations Deadline: Thursday, February 16th at 11:59 PM

Receive Prompt & Dataset: Friday, February 17th

Q&A Session and Presentation Tips: Friday, February 24th

Withdrawal Deadline: Monday, February 27th

Final Submissions (for 1st Round): Thursday, March 2nd by 11:59 PM EST

1st Round Presentations: Friday, March 3rd

Final Submissions (for 2nd Round): Thursday, March 9th by 11:59 PM EST

2nd Round Presentations/Travelers Day: Friday, March 10th

**Presentation Instructions / Judging Criteria:**

**First Round**

Your team will have 8 minutes to present your analysis and findings. Please think about the following questions to help guide your presentation. At the end of the presentation, you will have 5 minutes for Q&A with the judges.

* Describe the approach your team took to complete your data analysis.
* What kind of data analysis did you do?
* Do your visualizations tell a story or is it put together numbers?
* What type of story were you trying to convey with your analysis?
* Why did you choose to aggregate your data this way?
* Why did you choose the selected visualizations?
* What variables were most useful?
* Do you wish there were additional variables in the dataset? If so, what kind of variables?

**Final Round**

Your team will have 8 minutes to present your analysis and findings. Please think about the following questions to help guide your presentation. At the end of the presentation, you will have 10 minutes for Q&A with the judges.

**Judging – 1st and FinalRound:**

Each section will have an equal weight of 25% and will be judged using a structured rubric to ensure fairness.

* Data Analysis Findings
* Visualizations and Reasoning
* Documentation
* Final Group Presentation

**Final Submissions:**

Please email your university-specific contacts, listed below, by **11:59pm EST on Thursday, March 2nd** using the following format:

Subject your email with your “Team Name Final Submission” and attach:

* Data Analysis File
  + You can submit the code or the Excel notebook
  + Final Tables of your analysis
* Dashboard / Visualizations
  + Any charts, graphs, visualizations you created
* Documentation
  + Please provide the document as mentioned in the Documentation Section
* Final Presentation
  + Your Final Presentation Deck

**UConn Contacts:**

Julia Tremblay - [jetrembl@travelers.com](mailto:jetrembl@travelers.com)

McCullah Lamp - [mlamp@travelers.com](mailto:mlamp@travelers.com)

Jonathan Moore - [jonathan.a.moore@uconn.edu](mailto:jonathan.a.moore@uconn.edu)

**Eastern Contacts:**

Jafet Aparicio - [japarici@travelers.com](mailto:japarici@travelers.com)

Alex Citurs - [CitursA@easternct.edu](mailto:CitursA@easternct.edu)