Problem Descriptions

Problem 1 (submitted by Air Canada)
Contact Centre Staffing Forecasting

Context
Air Canada Cargo Customer Service Centres handle customer service inquiries for customers with freight originating, transiting or departing from Canada. We offer service through telephone and email and are reviewing alternative contact channels such as chat. Outside of standard customer service inquiries, the service centres handle new cargo bookings and specialty bookings such as Active Containers and Live Animals (horses, dogs, cats, etc.). The service centres in Toronto (YYZ), Montreal (YUL) and Moncton (YQM) are open from 0600 – 2200 EST, seven days a week and offer bilingual service to our customers.

Problem
With the ability to bid schedules only once per year, we need to forecast contacts and generate schedules that are linked across the three Customer Service Centres. Currently we rely on Excel spreadsheets and “heat mapping” to determine the heaviest contact volumes of the day and when staffing should be scheduled. This is only based on historical data, however, and takes into account neither language requirements (ENG/FR) nor future expected freight volumes.

Goal
The objective is to create a model that could estimate required staffing based on contact volumes taking into account our service level requirements and current staffing levels (# of total staff). This model should be based on an average estimated weekly.

Available Data
Here is the data that will be provided: any further required data will be discussed with the team.
1. The current staffing levels broken down by English/French qualifications.
2. Contact volumes by 30-minute increments for both telephone calls and emails as well as schedule samples.
3. Historical call volumes and wait times.
4. Historical cargo volumes and estimated future volumes.
5. Flight schedules.

Problem 2 (submitted by Air Canada)
Spot Rate Tool – Change Factor Simulator

Overview

Air Canada Cargo is an award-winning provider of air cargo transportation services. We are Canada's largest air cargo provider as measured by cargo capacity, with a presence in over 50 countries and self-handled hubs in Montreal, Toronto, Vancouver, Chicago, London, and Frankfurt. We connect over 450 cities across six continents with direct flights and hundreds more routes through our interline and trucking partnerships. Key investments in data and analytics, APIs, AI, and more, allow us to enhance our customer experience continually. One of those investments is an intuitive spot rate recommendation tool that will bring all relevant information to the user, while automatically providing a rate recommendation in an explainable and consistent manner.

Problem

The generation of Spot Rates is based on a set of business rules and input data that include base rates, revenue targets, and a rate change matrix that takes into account three elements: 1- Days to Departure, 2- Current Load Factor, and 3- Forecasted Load Factor. The change matrix will generate a "change factor" to increase or decrease the calculated spot rate. Currently the optimization of these "change factors" is mostly carried out manually.

Goal

The objective is to create a model/simulator that would allow us to test quickly and efficiently different combinations of change factors and evaluate their impact on revenues : this model/simulator would enable us to select a near-optimal change matrix.

Available Data

Here is the data that will be provided: any further required data will be discussed with the team.

1. Historical booking data.
2. Historical AWB data.
3. Market demand data.

Problem 3 (submitted by the Autorité des marchés financiers)
Data cleansing and transaction tracing

Cleansing of manually traced data

In the field of cryptocurrencies and DeFI (Decentralized Finance), transaction data are often complex and nested. They involve internal and external addresses, routers, intelligent contracts, and varied operations. Transactions including up to 12 internal transactions are not rare. As a rule an internal transaction consists of three swaps (token, counter-party, fee), with, for each swap, four transfers between the client, the target, a router, and a liquidity pool. Furthermore these transactions may use multiple paths, such as exchanges, conversions, multiple and direct transfers. The first step of the current project will consist of cleansing these data so that they can be used within an analytical or machine learning framework (see next paragraph). In this step transactions will be standardized and classified, and the originating transactions will be identified. To achieve this goal, we envision a global approach that considers all transactions of a given token, separates the reliable tracings from the unreliable ones, and then uses the reliable data to complete the unreliable data.

Tracing of transactions

In the first project, to extract a so-called complex transaction, it is necessary to reconstruct the price and reliable transactions around the transactions deemed unreliable. This approach makes tracing isolated transactions very costly. For this reason, a machine learning approach is considered. Once the data has been cleansed, a model can be trained to trace transactions automatically. This is where machine learning comes in. The aim is to develop a machine learning method capable of interpreting internal transactions to extract the true transaction. To achieve this, groups of previously cleansed data will be used to train a model whose inputs will be the internal transactions and the transaction metadata and whose output will be the amounts and names of the tokens exchanged. The architecture and approach of the model remain to be determined. It would be possible, for example, to divide and reconstitute the various internal swaps before carrying out the final tracing. It is important to note that reliability is a key factor in the success of the approach. A tracing reliability score could therefore represent a very significant added value.

Problem 4 (submitted by Environment and Climate Change Canada)
Estimate water level extremes at ungauged locations along the St. Lawrence fluvial estuary

Under the governmental Flood Hazard Identification and Mapping Program (FHIMP), Environment and Climate Change Canada (ECCC) has been mandated to provide 2D simulations of extreme water levels in the St. Lawrence fluvial estuary under historical and future conditions. The elevations of water levels in this system are triggered by the complex interaction of hydrological, meteorological, and tidal processes that must be considered to simulate river dynamics and flood events. Constraints on the computational resources and time requirements and the necessity for background geophysical fields currently limit the feasibility of producing fine-scale 2D hydrodynamic simulations to a limited set of relatively short extreme events (approximately 400 events with durations ranging from one hour to several weeks).

Hence several complementary modelling tools have been explored to study the temporal evolution of water level extreme properties. Among them, some multivariate statistical models and Machine Learning (ML) tools have proven effective in reconstructing continuous water level series over long historical periods, which is essential to assess the extreme probability distributions. However, while these methodologies have shown promising performance at gauged stations, the challenge remains in extending their applicability to other ungauged locations within the estuary where only short 2D simulations are available. During the workshop, we are thus considering two specific questions:
1. How can we estimate the extreme characteristics (e.g., return period, duration, and seasonality) at ungauged locations by leveraging 2D short hydrodynamics simulations run for extreme events (few days to few weeks) and long water-level reconstructions obtained at some locations (e.g., gauge stations) using the statistical or ML tool?
2. How do we assess the reliability of the hourly reconstructions and extreme estimates at ungauged locations to determine the most suitable strategy for implementation in the FHIMP project?

ECCC will provide data suitable for a benchmark analysis, including the following datasets for the 1970-2022 period: hourly water level records observed at 15 stations over the study domain; hourly water level reconstructions at 2 stations obtained with a non-stationary tidal harmonic regression tool and the corresponding regressors; 2D hydrodynamics simulations corresponding to a subset of extreme events observed at the selected stations.

Problem 5 (submitted by the International Air Transport Association)
Characteristics and forecasting of turbulence

Turbulence is a major cause of non-fatal injuries in the aviation industry and costs millions of dollars every year. To help pilots avoid turbulence, reduce injuries, and lower emissions, the International Air Transport Association (IATA) has created the largest live-turbulence data platform in the industry.

As part of IATA's continuous effort to improve safety, we want to analyze the data we have collected since 2019 and gain valuable insights. Turbulence is a volatile phenomenon. The first question we want to address is one that meteorologists have been attempting to answer for a long time: How long does turbulence last? And does its duration depend on the time of the year? Is there a relationship between turbulence duration and altitude?

The answer to the above question is the basis for answering the second question we want to address. Pilots can be alerted if the airplane's trajectory points toward a turbulent area. Given the aircraft's location, heading, and speed, we want to determine the likelihood of turbulence ahead of that aircraft based on live and historical turbulence data. Such information can then be dispatched to the pilots in real-time so they can make informed tactical decisions and avoid turbulence.

We want to answer one last question about thermal-based turbulence, which causes significant "bumps" during the flight descent phase. That phase of the flight is important as flight attendants are usually securing the cabin during these times. We want to count the number of thermal-based turbulent events and their frequency based on EDR and cloud cover data at low levels.

Problem 6 (submitted by Revenu Québec)
Pattern identification in a transactions data bank

Asset transactions, or sequences of transactions, that occur between different parties, whether individuals or companies, may be accompanied by tax obligations that vary according to the field of application. In some cases, certain parties form groups and develop transactional strategies to avoid their tax obligations. In order to fulfill its mission of ensuring tax compliance, Revenu Québec is keen to uncover emerging behaviours in the context of asset transactions.

With this in mind, one could use an approach based on the identification of transaction patterns. Once identified, these patterns could be analyzed by experts in the business field to determine whether they really represent a pattern of behaviour designed by a group of players with the aim of evading their tax obligations.

To identify a network, a bottom-up approach can be used, starting with the smallest details and gradually discovering more complex structures. In other words, the identification of this kind of non-compliant behaviour can begin with the detection of a suspicious participant or transaction. Subsequently, information on other participants and transactions is sought, and a strategy for evading tax obligations may be uncovered.

In the present context, we would like to take a top-down approach to the problem, i.e., consider all the transactions, participants, and links, and attempt to identify patterns that would lead us to groups of participants and transactions that merit further analysis in the search for non-compliant behaviours.

The data provided will be anonymized transaction data containing at least the following informations.
1. Participants : their types, links with other participants, and some other relevant characteristics.
2. Assets : their types and some relevant characteristics.
3. Transactions : their types and dates and the assets and participants involved.

From such data one could identify frequent patterns or association rules that would enable the discovery of strategies to be analyzed. One could also detect anomalies in the structure of relationships, or use a graph representation to detect interesting subgraphs. We would prefer a search based solely on the structure of relationships and not on previously uncovered offender networks.

This approach is completely new at Revenu Québec. Thus our expectations and results are not defined in a precise fashion. In particular such concepts as ‘anomaly,’ ‘interesting’ or ‘relevant’ are not defined formally in a business context. The main goal of this project is to explore the potential of decontextualized transaction data for the automatic discovery of networks suspected of malicious behaviour. We expect that tools from data science will enable one to identify some configurations that are abnormal with respect to the provided data set. Furthermore one should be able to explain and reproduce the method for identifying such configurations.