Additional information for the creation of case studies

Course: Case Study: Model Engineering (DLMDSME01)

Table of contents

1. Data-Set: Credit Card Routing for Online Purchase via Predictive Modelling

1.1 File Name:

- zip-folder: use_case_1.zip
- *PSP_Jan_Feb_2019.xlsx* : list of credit card transactions for DACH countries (Germany, Switzerland, Austria)

1.2 List of PSPs (=payments service providers) and service fees:

name	Fee on successful transactions	Fee on failed transactions
Moneycard	5 Euro	2 Euro
Goldcard	10 Euro	5 Euro
UK_Card	3 Euro	1 Euro
Simplecard	1 Euro	0,5 Euro

1.1 Column Description:

- tmsp: timestamp of transaction
- country: country of transaction
- amount: transaction amount
- success: is 1 if payment is successful
- PSP: name of payments service provider
- 3D_secured: is 1 if customer is 3D identified (i.e. more secure online credit card payments)
- card: credit card provider (Master, Visa, Diners)

1.2 Additional Info from Business Side:

- Many transactions fail at the first try. Therefore, customers try several times to transfer the money. If two transactions are within one minute, with the same amount of money and from the same country, it is (for a decent number of tries) safe to assume that they are payment attempts of the same purchase. Consider this possibility of several payment attempts of the same purchase in your machine learning model!

2. Automation of Standby Duty Planning for Rescue Drivers via a Forecasting Model

2.1 File Name:

- zip-folder: use_case_2.zip
- *sickness_table.csv*: daily information on sickness counts, emergency calls, available standby resources and how many additional resources are activated

2.2 Column Description:

- date: entry date
- n_sick: number of drivers called sick on duty
- calls: number of emergency calls

- n_duty: number of drivers on duty available
- n_sby: number of standby resources available
- sby_need: number of standbys, which are activated on a given day
- dafted: number of additional drivers needed due to not enough standbys

2.3 Additional Info from Business Side:

- Business claims, that having a daily fixed amount of standbys (n_sby = 90) is not efficient because there are days with too many standbys followed by days with not enough standbys. The business aims at a more dynamical standby allocation, which takes seasonal patterns into account.
- Most important, the model should minimize dates with not enough standby drivers at hand!

3. Efficient Flight Operations at East Carmen Airlines via a Machine Learning Flight Prediction Model

3.1 File Names

- zip-folder: use_case_3.zip
- flight_information.csv : information on flights in a given time period, where each data point is a so-called leg, i.e. a flight from departure airport to destination airport with all relevant flight and crew information
- ground_information.csv : information on the ground processes after landing at our hub "East Carmen", where each data point describes the processes between two flights (inbound = landing, outbound = take-off).

3.2 Column Description – Flight Information

- *leg_no*: unique identifier of a flight on a given day, at a certain time, with a given flight number
- *fn_carrier*: airline name
- *fn_number*: flight number; has to be unique on a given day
- *dep_ap_sched*: scheduled departure airport
- *arr_ap_sched*: scheduled arrival airport
- *dep_sched_date*: scheduled departure date
- *dep_sched_time*: scheduled departure time
- *arr_sched_date*: scheduled arrival date
- *arr_sched_time*: scheduled arrival time
- *m_offblockdt*: timestamp of departure
- *m_onblockdt*: timestamp of arrival
- *ac_registration*: aircraft registration number, i.e. the "license plate" of the aircraft
- *change_reason_code*: reason for delay (assigned after the flight)
- *dep_delay*: departure delay
- *Ac Type Code*: aircraft type (example: 320 = Airbus A320)
- *trans_time*: true minimal transition time for crew members after flight, i.e. transition time of the crew member with the least time
- *sched_trans_time*: scheduled minimal transition time for crew members after flight, i.e. scheduled transition time of the crew member with the least time

- *Crew Group*: assignment of what happens to the whole crew after a flight
  - Start : First flight of day
  - A : all crew members stay on the aircraft for the next flight
  - B, B2: all crew members switch aircraft for the next flight
  - C: at least one crew member switches aircraft for the next flight
- *TLC_trans*: names of crew members on flight with some additional information attached to each name (but business does not know exactly which additional information is visible there)
- *crew_type_change*: rank (cp = pilot, ca = cabin member) of crew members, who changed aircraft
- *Sched Groundtime*: scheduled ground time of the aircraft between flights
- *Act Groundtime*: actual ground time of the aircraft between flights

3.3 Column Description – Ground Information

- Each datapoint consists of an inbound (=arriving) flight and an outbound (=departing) flight from our hub "East Carmen)
- Catering, cleaning and pax (=passenger) boarding durations given for our hub "East Carmen" between the inbound and outbound flights
- *day_of_origin*: day of flight (also given in Flight Information)
- *ac_type*: aircraft type (also given in Flight Information)
- *fn_number*: flight number (also given in Flight Information)
- *ac_registration*: aircraft registration (also given in Flight Information)
- *mingt*: minimal scheduled ground time for the given aircraft
- *dep_leg_inbound*: departure airport name of inbound (=arriving) flight
- *arr_leg_inbound*: arrival airport name of inbound (=arriving) flight
- *arr_leg_outbound*: departure airport name of outbound (=departing) flight
- *sched_inbound_dep*: scheduled departure time of inbound (=arriving) flight
- *sched_inbound_arr*: scheduled arrival time of inbound (=arriving) flight
- *sched_outbound_dep*: scheduled departure time of outbound (=departing) flight
- *sched_outbound_arr*: scheduled arrival time of outbound (=departing) flight
- *sched_turnaround*: scheduled ground time for aircraft
- *leg_inbound*: leg number of inbound flight
- *leg_outbound*: leg number of outbound flight
- *catering_duration*: catering duration (i.e. filling up meal boxes) between flights in minutes
- *cleaning_duration*: cleaning duration between flights in minutes
- *pax_boarding_duration*: boarding duration between flights in minutes

3.4 Additional Info from Business Side:

- BI (=business intelligence) colleagues have collected a lot of flight information from many different departments: flight-, crew- and ground operations. Each department uses a different data warehouse. BI colleagues claim that some of the columns might be redundant and some information could be wrong, due to bad and not consistent data in the different data warehouses of the business departments. It is central for this prototype study to clean and transform the data and make consistency checks, before applying a machine learning algorithm.