TABLE OF CONTENTS

AGENDA OF EVENTS ........................................... 3

REVIEW: MACHINE INTELLIGENCE SUMMIT 2017 ........................................... 3

DESIGNS OF EXPERIMENTS IN THE CONTEXT OF SMALL DATA ......................... 4

TIME SERIES ANALYSIS ON BIG DATA ........................................... 6

APPLYING TIME SERIES ANALYSIS ON AIRCRAFT TRANSPONDER DATA ............. 13

A PLATFORM TO FOSTER MULTIDISCIPLINARY DATA SCIENCE COLLABORATIONS .......... 18

ANALYTICS IN BANKING ........................................... 20

WEAPONS OF MATH DESTRUCTION ........................................... 22

SUPPORT THE SWISS ASSOCIATION FOR ANALYTICS ! ........................................... 23

DATA SCIENCE: BEYOND ALL THE HYPE, WHAT YOU NEED TO KNOW IN A FEW ARTICLES, BUZZ-FREE!

Nowadays, more than ever, our field is under the spotlight as people are increasingly talking about Artificial Intelligence, Machine Learning and Analytics in general. This is why we think that going back to the basics with essential topics depicted in this magazine is important. Our goal has always been to accompany you in your analytics journey and we hope that our selected articles will provide you with useful insights.

In this release, we will start by having a feedback of the machine intelligence summit by Hanne Vlaeminck, then we’ll have Nicolas Antille explaining the very valuable design of experiments in the context of R&D. Afterwards, moving on to Big Data, Marco Gessner and Maurizio Felici will walk us through the equally known and important field of Time Series, but here applied with large volumes of data and Mark Whalley shows us an applied case on aircraft transponder data. Stan Schymanski will then present the Swiss Data Science Center and their RENGA platform. Subsequently one cannot forget that institutions that are present in our country and rely widely on data are banks. In this sense, Nathalie Feingold presents us what lies within analytics in banking. Additionally, Sandro Saitta reviewed the book “Weapons of Math Destruction”.

I take the opportunity to remind you of our main platforms of communication:
- Meetup (www.meetup.com/swiss-analytics):
  Analytics events proposed by our association.
- LinkedIn (www.linkedin.com/groups/4586163):
  Our discussion group (news, job offers, events, etc.)
- Website (www.swiss-analytics.com):
  Content we share (magazine, event slides, etc.)

Our next General Assembly is approaching, and we hope to be able to count on you. We would like to remind you that to become a free member, you need to simply join our LinkedIn group. Until we meet again, the whole Committee would like to wish you great end of year holidays.

Amine Mansour, Big Data and Data Science Leader at Itecor
President of the Swiss Association for Analytics
REVIEW: MACHINE INTELLIGENCE SUMMIT 2017

The Machine Learning Summit is an event that aspires to bring together government officials, academics, entrepreneurs, investors, and industry professionals all in the same room to discuss the latest breakthroughs in machine learning, deep learning and AI and to explore their applications in industry and their potential impact...

I attended the 4th Machine Intelligence Summit that took place in Amsterdam 28-29 June 2017. There were over 30 speakers and 200 attendees. From a hand-raising survey of the room on the 2nd day, it looked like the audience was mostly from an academic or startup background, with a good number of industry professionals as well, both in the audience as well as speaker.

Opportunities

This Machine Intelligence Summit was an interesting conference. The way the venue is set up and the overall organization really encourages networking and meeting a mix of diverse and interesting people. Many industry professionals clearly were there looking to find the next interesting machine learning application for their company.

It’s no surprise that overall there was a lot of excitement about deep learning, though there were also interesting applications using more “old school” machine learning approaches. Several presenters illustrated the tremendous progress that has been made in a short time-span regarding image recognition, segmentation and object detection. In this domain there are now possibilities that did not exist just a couple of years ago.

Challenges

The importance of having relevant business problems, the right data, and realistic expectations for successful machine learning applications were stressed several times throughout the conference. Many of the speakers had slides about the challenges that they currently face in their machine learning and deep learning projects, spanning the entire product life cycle. Examples of such challenges include:

• Difficult communication and expectation management between data scientists and product managers.
• Deep learning is extremely data-hungry, which limits the possible applications.
• “The model does not know what it does not know” and it will not warn you when not to trust its predictions which limits its use in mission critical applications.
• State of the art AI is not automated: there is still a lot of human involvement required.

From having the answers to posing the right questions

One very good point made by one of the speakers is that the application of data and machine learning to problem spaces that traditionally relied on domain knowledge is shifting the definition of ‘expertise’ from having the answers to known questions towards knowing to ask good questions [and whether they can be answered with data].

Several other speakers put the emphasis on the fact that, while deep learning is highly fashionable right now, the first focus should be on solving relevant business problems, make sure you have the right data, and let the “fancy” come later.
DESIGNS OF EXPERIMENTS IN THE CONTEXT OF SMALL DATA

It’s not all about Big Data

There is no doubt that Big Data has become a buzz word. For very good reasons since the development and deployment of high-throughput data acquisition systems facilitates a lot the collection of large and diverse datasets. There are numerous examples in the context of the food and beverages industry with the development of apps to capture consumption habits and eating behaviors, of wearable devices to record physiological data and of sensors to measure product characteristics online during the manufacturing process. These new acquisition systems represent great opportunities to progress towards a more personalized food offering or towards self-automated production lines, but they should not let us forget that R&D remains today above all a world dominated by Small Data.

Experimentation is costly

At Nestlé, experimentation remains the most straightforward approach to collect data in the context of product innovation or renovation projects, and for many categories every single trial requires a non-negligible amount of raw material, time and money. As an example, producing a yogurt in a pilot plant takes more than one day because of the time needed for fermenting the product, whereas it costs up to 50’000 CHF to produce one sample of soluble coffee (still at the pilot scale), in particular because of the complexity and depreciation of the equipment and of the number of operators required at each process step. As resources are far from being unlimited (unfortunately...), it means that the number of trials that can be executed in the frame of a particular study or project is often less than 20. We are then quite far from the Big Data world!

Optimizing the experimental effort

As statisticians contributing to these projects we are on the front line to propose solutions that allow to achieve project objectives while meeting constraints, for instance in terms of pilot plant availability or sensory analysis capabilities. Luckily, we have developed over the years a strong expertise in Design of Experiments (DoE) and we can rely on a full range of approaches particularly adapted to the food and beverages context. We have even developed our own strategies¹ to deal with complex situations for which no satisfactory solution could be found in the literature or in dedicated software.

Most people using DoE claim as the top benefit that it helps reducing the experimental effort, but at Nestlé we rather claim that it helps optimizing the experimental effort. Indeed, most of the time the size and complexity of the problem require to use all available resources (e.g. pilot plant slots) to solve it, and an experimental strategy based on DoE helps maximizing the information obtained from performed trials. In this context, we work closely with product and process experts to define the right set of factors (and associated levels) and identify possible constraints upfront. This contributes to take already existing knowledge into account when designing the trials, to reduce the risk to run non-relevant trials and to come up in the end with a set of fit-for-purpose trials.

¹ A. Rytz et al., Using fractional factorial designs with mixture constraints to improve nutritional value and sensory properties of processed food, Food Quality and Preference, Volume 58, June 2017, Pages 71-75
A step towards the digital twin

Another key benefit of DoE is that it creates the ideal conditions to build causal descriptive models, which is essential at Nestlé since the expected outcome of most product (or technology) development studies is a quantification of the impact of some input variables (e.g. ingredients, process parameters) on product characteristics (e.g. sensory, instrumental). Using DoE strategies, in particular those based on orthogonality, enables to strongly limit collinearities between input variables and so to more easily isolate the contribution of each to differences observed between products. We then often translate the outcome of this statistical modelling work into formulation guidance tools (Excel or Shiny format) that product developers use to predict characteristics of any product within the studied product space or to identify cost-effective, nutritionally-balanced or even ecologically-friendly products matching a target sensory profile (digital twin approach). All this contributes to accelerate the development and launch of products meeting consumer needs.

A methodology rather than a tool

Overall, we like to position DoE as a knowledge-building or problem-solving methodology (and not just as a mathematical tool) because:

1. It helps scoping the project and aligning the project team (often cross-functional) around clear objectives.
2. It helps performing trials that are specifically designed to achieve defined objectives.
3. It helps building causal models contributing to make knowledge sustainable and reusable.

The DoE journey started more than 25 years ago at Nestlé, and DoE is now widely used across the whole R&D organization. The numerous awareness sessions delivered by statisticians of course contributed to this success, but the main success factor remains the fact that the DoE approach demonstrated a tangible added value in hundreds of projects.

Outlook

As Big Data seems to have a quite big potential for engaging the dialog with consumers (for instance through Apps) about their preference, eating habits and nutrition concerns, or for checking and ensuring products safety and quality, Small Data will likely remain the norm for most fundamental research or product development projects in the coming years. This implicitly means that Design of Experiments has still a long life ahead at Nestlé R&D. However, new DoE-based strategies may be required at some point to deal with new experimentation approaches, like fast prototyping.
TIME SERIES ANALYSIS ON BIG DATA

Marco Gessner, Senior Architect at Vertica

Introduction

Time series refer to an ordered sequence of values of a variable at equally spaced time intervals. Many phenomena in our daily lives are measured in intervals over a period of time. Examples of time series range from corporate business metrics like weekly sales to stock prices to sensors’ measures to monitoring industrial processes.

As the world gets more and more connected, instrumented and intelligent, sensors become ubiquitous collecting tens of thousands of measurements every seconds thus leading to an enormous amount of data being produced on a daily basis.

In order to get insights from these ‘large’ time series and make informed decision it is important to have the ability to store and manipulate it in an efficient manner.

In this article we will review time series analysis and discuss what it entails from a data preparation perspective. We will then highlight challenges introduced by the IoT wave and provide requirements for an analytics platform to analyze large scale time series.

Time series applications

What we can do with a time series? Here you have an (incomplete) list of very common time series Analysis:

1. **Forecasting.** We want to predict future values for the measure based on the previous ones. For example we might want to forecast electricity demand based on previous consumption data

2. **Cross-Correlation.** This analysis is used to discover if/how changes in one time series influence the behavior of other time series. For example how temperature affects electricity demand or... ice creams sales. How stock price changes of two different companies are related each other.

3. **Auto-Correlation.** This is a special case of Cross-Correlation where we have in input a single time series correlated with itself at a different time to discover “signal periodicities”. For example: temperature in a given location will probably show daily and yearly seasonalities

4. **Similarity** (Pattern Matching). This is used to measure the “distance” between two (or more) time series. Think for example about ECG waveforms: we might want to check if a 10 second long ”shape” already happened in the past.

5. **Anomaly Detection.** Anomaly detection is the ability to automatically discover statistical anomalies in the signal based on the previous values and – for example – trigger an alarm. You might want to consider “suspicious” signal values that are – let’ say – 3 standard deviations above/under its average value in the last X seconds.

Time series analysis prerequisites

Many time series analysis require evenly distributed time series. Meaning that the interval between two consecutive measures has to be constant and – if we compare multiple time series – exactly the same. However, in the real life, irregularly distributed time series are quite common:

- Some sensors, to save bandwidth and/or conserve battery, send their measures only when they notice a change. For example, if one of these “on-change-only” devices is used to measure the
temperature you might receive one data point 17°C at 09:41 and another measure 18°C at 10:25. This means that the temperature was still 17°C anytime between 09:41 and 10:25.

- In other cases, data can just be unavailable from the source. Think for example about astronomic observations from the Earth when weather conditions are not good enough.

- Or... measures can just be lost after their acquisition.

So... transforming these irregularly distributed time-series into regular ones is one of the first steps to do before starting analyzing your time series.

Rebuilding these artificial data points requires both:

- **Gap Filling**: rebuilding missing points on the time axis with a regular interval.

- **Interpolation**: assign a value to the missing data point. Different Interpolation techniques can be used to define missing values. For “on-change-only” sensors we just keep using the last known value /constant interpolation).

Then we have to consider prerequisites specific to the Time Series Analysis we want to run. For example, in order to run an ARIMA forecasting model, we have to verify if our Time Series is stationary. This means that statistical properties like mean, variance and autocorrelation do not change (too much) over time. And, if this is not the case, “transform” the Time Series before running the analysis. So, for example, if you have a time series like this:

You might need to detrend this time series before running ARIMA. The number and type of these propaedeutic transformation depends on the analysis we have in mind. Common transformation include:

- **Detrending.** The example shown here above is a (very) simple deterministic (and linear) trend which is quite easy to recognize and remove but there are other situations where the “trend component” is neither deterministic nor linear. In this cases detrending is not as easy as removing a straight line.

- **Smoothing/denoising.** Used to remove irregular and unpredictable components from a signal. Sometimes these components are considered as unwanted “noise”. There are dozens of different algorithms one could implement in this case. A very common and easy to implement method is based on simple moving averages (where all data points have the same weight) or weighted moving averages (where older data points have a smaller weight). A special case is exponential smoothing where weights are exponentially decreasing.

- **Decomposition.** Decomposition is used to split time series into their trend, seasonal and irregular components. To give you an example here.

```r
Source code (R):
> time <- rep(1:100)  # Create time vector
> var <- rnorm(100)    # Generate random noise
> value <- time * 0.1 + var
> plot(time, value, type='l')
> abline(lm(value~time), col='blue')
```

![Graph showing detrending example](image-url)
we decompose air passenger data between Jan-1949 and Dec-1960:

- **Offset Translation.** This is normally used to make it easier comparing time series having different “offsets” (additive component) on their value axis

- **Amplitude Scaling.** This technique, similarly to offset translation, is used to make time series uniform along the value axis but in this case we consider the multiplicative (not additive) terms

- **Spectrum Analysis.** This technique is used to identify how “strong” are the different spectral density components in our time series and is normally implemented using Fast Fourier Transform.

**Big Data and Time Series Analysis**

As above mentioned the IoT wave introduces new challenges in terms of data storage and analysis. We argue that that an analytics platform should exhibit the following feature to efficiently store and analyze time series at a large scale.

1. **The ability to ingest data at very high speed.** Even if time series representation is normally quite simple (SENSOR_ID, TIMESTAMP, VALUE) the volumes generated by IoT data sources (smart meters, web logs, satellites, medical data,...) are huge

2. **The availability of basic transformations** like for example Gap Filling and Interpolation to convert irregular time series into regular ones and/or the possibility to “join” time series with different and irregular time axis

3. **The possibility to implement ad-hoc transformations and analysis “inside” the database** using advanced statistical-oriented programming languages. Analyzing data outside the database will oblige you to move huge amounts of data back and forth and this will introduce huge latencies. Several databases offer the possibility to deploy and run code “inside” the database in order to run transformations as close as possible to the data.

**Manipulating Time Series with Vertica**

Vertica perfectly lends itself to times series storage and analysis. Vertica’s encoding capability allow for optimal representation of time series, specifically by using Run Length Encoding and appropriate sorting it is possible to drastically reduce the footprint on disk and at the same time accelerate queries response time. This is mainly due to the fact that timestamp and measure_id, two typical columns in time series have a low cardinality with respect to the total number of measures.

From an analytics perspective there are a number of predefined functions and specific clauses in Vertica SQL that efficiently help in the data preparation process, and we will describe some of them in the next paragraphs.

**Source code (R):**
```r
> library(Ecdat)
> data(AirPassengers)
> fit=stl(AirPassengers, s.window="periodic")
> plot (fit)
```
These are best explained with a small input data extract, the code of the transformation, and the results of the transformation. They are obviously best understood with an initial knowledge of SQL.

**Gap Filling And Interpolation**

The example data are electricity smart meter readings. They often come with long gaps, as sometimes the electricity companies use their electric power lines as communication lines to transmit the smart meters’ readings to the collecting central, and the high-voltage transmission lines sometimes need to be disconnected to keep a grid in balance.

The time series here consists of the identifier of the smart meter, the reading timestamp, and the running sum of consumed Watt hours delivered through this delivery point up to now:

<table>
<thead>
<tr>
<th>meter_id</th>
<th>reading_ts</th>
<th>watt_hrs</th>
</tr>
</thead>
<tbody>
<tr>
<td>100,012,700,004,100</td>
<td>2015-01-01 03:00:00</td>
<td>10,000</td>
</tr>
<tr>
<td>100,012,700,004,100</td>
<td>2015-01-01 03:03:00</td>
<td>10,030</td>
</tr>
<tr>
<td>100,012,700,004,100</td>
<td>2015-01-01 03:20:00</td>
<td>10,370</td>
</tr>
</tbody>
</table>

Note that the time series is irregular; a first measure on the hour, one after three minutes, and one after another seventeen minutes. Also note that, in 3 minutes, 30 Watt hours were consumed: 10 Wh/min; and in the following 17 minutes, we see 340 Wh consumed: 20 Wh/min.

If this is a scenario where we want to forecast electricity consumption, using linear regression, we require an evenly spaced time series, for which we choose a regular interval of 5 min, and we need a linear, not a constant, interpolation of the Wh values. While in classic SQL this requires a table with evenly spaced timestamps, joined in a complex manner with the base table, and extensive date arithmetic to calculate the linear interpolation, Vertica lets us do this with a two line clause, plus a function depending on it:

```sql
SELECT
  meter_id, interval_5min::TIMESTAMP(0),
  ts_first_value(watt_hrs, 'linear') AS watt_hrs_5min
FROM energy_readings
TIMESERIES interval_5min AS '5 MINUTES' OVER
(PARTITION BY meter_id ORDER BY reading_ts)
```

This is the result:

<table>
<thead>
<tr>
<th>meter_id</th>
<th>interval_5min</th>
<th>watt_hrs_5min</th>
</tr>
</thead>
<tbody>
<tr>
<td>100,012,700,004,100</td>
<td>2015-01-01 03:00:00</td>
<td>10,000</td>
</tr>
<tr>
<td>100,012,700,004,100</td>
<td>2015-01-01 03:05:00</td>
<td>10,070</td>
</tr>
<tr>
<td>100,012,700,004,100</td>
<td>2015-01-01 03:10:00</td>
<td>10,170</td>
</tr>
<tr>
<td>100,012,700,004,100</td>
<td>2015-01-01 03:15:00</td>
<td>10,270</td>
</tr>
<tr>
<td>100,012,700,004,100</td>
<td>2015-01-01 03:20:00</td>
<td>10,370</td>
</tr>
</tbody>
</table>

Note that the measure three minutes after the hour has disappeared; that the difference of 70 Wh between the first and second row consists of 10 Wh/min for 3 minutes and 20 Wh/min for 2 minutes, and the remaining rows have a regular difference between each other of 20 Wh/min for 5 minutes.

**Join Time Series with Irregular and Different Time Axis**

This can be illustrated with two sets of measures in a combustion engine, oil pressure and engine rotational speed:

<table>
<thead>
<tr>
<th>op_ts</th>
<th>op_psi</th>
<th>rs_ts</th>
<th>rpm</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015-04-01 07:00:00</td>
<td>25.356</td>
<td>2015-04-01 07:00:01</td>
<td>25.201</td>
</tr>
<tr>
<td>2015-04-01 07:00:10</td>
<td>35.124</td>
<td>2015-04-01 07:00:08</td>
<td>3.508</td>
</tr>
<tr>
<td>2015-04-01 07:00:20</td>
<td>47.056</td>
<td>2015-04-01 07:00:15</td>
<td>6.504</td>
</tr>
<tr>
<td>2015-04-01 07:00:30</td>
<td>45.225</td>
<td>2015-04-01 07:00:20</td>
<td>6.608</td>
</tr>
</tbody>
</table>
Only 2 rows, in the two tables, can join over an equal timestamp:

At 2015-04-01 07:00:00 and 2015-04-01 07:00:20.

A full outer equi-join would produce this report:

<table>
<thead>
<tr>
<th>op_ts</th>
<th>op_psi</th>
<th>rs_ts</th>
<th>rpm</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015-04-01 07:00:00</td>
<td>25.356</td>
<td>2015-04-01 07:00:00</td>
<td>2,201</td>
</tr>
<tr>
<td>2015-04-01 07:00:10</td>
<td>35.124</td>
<td>2015-04-01 07:00:15</td>
<td>16,504</td>
</tr>
<tr>
<td>2015-04-01 07:00:20</td>
<td>47.056</td>
<td>2015-04-01 07:00:20</td>
<td>6,608</td>
</tr>
</tbody>
</table>

While classic SQL would require first the use, on each of the input tables, of the LEAD() OLAP function, to obtain the timestamp immediately following in the time series, followed by a complex range join using the original and the derived timestamps, Vertica offers the INTERPOLATE PREVIOUS VALUE predicate, to be used instead of the equal operator:

```
SELECT * FROM oilpressure
FULL OUTER JOIN revspeed
ON op_ts INTERPOLATE PREVIOUS VALUE rs_ts;
```

Its effect is that each row of one table is joined with the row of the other table immediately preceding this row in time:

<table>
<thead>
<tr>
<th>op_ts</th>
<th>op_psi</th>
<th>rs_ts</th>
<th>rpm</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015-04-01 07:00:00</td>
<td>25.356</td>
<td>2015-04-01 07:00:08</td>
<td>3,508</td>
</tr>
<tr>
<td>2015-04-01 07:00:10</td>
<td>35.124</td>
<td>2015-04-01 07:00:15</td>
<td>16,504</td>
</tr>
<tr>
<td>2015-04-01 07:00:20</td>
<td>47.056</td>
<td>2015-04-01 07:00:20</td>
<td>6,608</td>
</tr>
</tbody>
</table>

With that, we are now able to correlate engine rotational speed to oil pressure in one row.

Separating Subsets of Time Series into “Sessions”

A repeatedly occurring requirement in time series is to be able to stop analyzing a time series when it is interrupted for a certain period of time. For example, an aircraft could stop emitting signals once the engine is shut off, and start again once the engine has restarted, as can be seen in the series that follows, where we have a gap of 1 hr 1 min between two rows:

<table>
<thead>
<tr>
<th>flightno</th>
<th>ts</th>
<th>fuel_ltr</th>
</tr>
</thead>
<tbody>
<tr>
<td>SR450</td>
<td>2017-11-12 15:23:00</td>
<td>51,827</td>
</tr>
<tr>
<td>SR450</td>
<td>2017-11-12 15:24:00</td>
<td>51,826</td>
</tr>
<tr>
<td>SR450</td>
<td>2017-11-12 15:25:00</td>
<td>51,825</td>
</tr>
<tr>
<td>SR450</td>
<td>2017-11-13 15:26:00</td>
<td>183,380</td>
</tr>
<tr>
<td>SR450</td>
<td>2017-11-13 15:27:00</td>
<td>183,370</td>
</tr>
<tr>
<td>SR450</td>
<td>2017-11-13 15:28:00</td>
<td>183,365</td>
</tr>
</tbody>
</table>

Vertica offers its own CONDITIONAL_TRUE_EVENT() OLAP function; it returns an integer that is incremented every time an expression passed to the function is true. In this specific case, that would be that the difference between the timestamp of the preceding row (obtained with the LAG() OLAP function) and the timestamp of the current row exceeds half an hour:

```
SELECT
  CONDITIONAL_TRUE_EVENT(ts - LAG(ts) > '30 minutes')
OVER (PARTITION BY flightno ORDER BY ts)
  AS sess_id
, flightno
, ts
, temp
FROM sdata;
```

```
sess_id|flightno |ts            |fuel_ltr |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>SR450</td>
<td>2017-11-12 15:23:00</td>
</tr>
<tr>
<td>0</td>
<td>SR450</td>
<td>2017-11-12 15:24:00</td>
</tr>
<tr>
<td>0</td>
<td>SR450</td>
<td>2017-11-12 15:25:00</td>
</tr>
<tr>
<td>1</td>
<td>SR450</td>
<td>2017-11-13 15:26:00</td>
</tr>
<tr>
<td>1</td>
<td>SR450</td>
<td>2017-11-13 15:27:00</td>
</tr>
<tr>
<td>1</td>
<td>SR450</td>
<td>2017-11-13 15:28:00</td>
</tr>
</tbody>
</table>
```
This allows the analysis of separate phases of a time series.

Event Series Pattern Matching: Filter a Number of Sub Time Series That Match a Defined Behavioral Pattern.

This is also a recurring requirement in time series: the need to filter out, from a base time series, sequences of rows that correspond to a defined behavioral pattern, combined with the need to identify each found behavioral pattern differently.

The following table contains 30 rows, coming from two test drives of an experimental car, with measures for engine rotational speed (rpm) and airmass intake per unit of time (airmass).

<table>
<thead>
<tr>
<th>test_drive_id</th>
<th>vehicleid</th>
<th>ts</th>
<th>rpm</th>
<th>airmass</th>
</tr>
</thead>
<tbody>
<tr>
<td>27</td>
<td>3</td>
<td>00:00:00</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>27</td>
<td>3</td>
<td>00:00:10</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>27</td>
<td>3</td>
<td>00:00:20</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>27</td>
<td>3</td>
<td>00:00:30</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>27</td>
<td>3</td>
<td>00:00:40</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>27</td>
<td>3</td>
<td>00:00:50</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>27</td>
<td>3</td>
<td>00:01:00</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>27</td>
<td>3</td>
<td>00:01:12</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>27</td>
<td>3</td>
<td>00:01:23</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>27</td>
<td>3</td>
<td>00:01:33</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

27 | 00:05:10 | 300 | 5
27 | 00:05:20 | 400 | 7
27 | 00:05:30 | 430 | 7
27 | 00:06:00 | 340 | 4
27 | 00:06:12 | 40  | 1
27 | 00:06:23 | 0   | 0
29 | 00:00:08 | 0   | 0
29 | 00:00:17 | 5   | 0
29 | 00:00:26 | 6   | 0
29 | 00:00:37 | 7   | 0
29 | 00:00:48 | 4   | 0
29 | 00:01:12 | 1   | 0
29 | 00:01:23 | 0   | 0
29 | 00:01:33 | 0   | 0
29 | 00:02:51 | 300 | 5
29 | 00:03:01 | 400 | 6
29 | 00:03:17 | 430 | 7
29 | 00:03:25 | 340 | 4
29 | 00:03:43 | 40  | 1
29 | 00:03:57 | 0   | 0

From this table, the requirement is to analyze the development of air mass intake in all failed / terminated engine start attempts. A failed start attempt can be recognized by the fact that, between two moments when the engine does not turn (rpm = 0), the engine never reaches the tick over speed of 750 rpm. In the data above, the three highlighted sub-series satisfy that condition: once during test drive # 27, twice during test drive # 29.

While it’s highly probable that this requirement can’t be solved using classic SQL at all – Vertica offers the MATCH() clause, which allows to specify a series of events in a satisfyingly intuitive manner:

```sql
WITH
  with_prev_rpm AS (
    SELECT
      LAG(rpm) OVER(PARTITION BY test_drive_id ORDER BY ts) AS prev_rpm,
      * FROM sdata
  ),
  interesting_rows AS (
    SELECT
      * , event_name()
      , pattern_id()
      , match_id()
    FROM with_prev_rpm
    MATCH ( 
      PARTITION BY test_drive_id 
      ORDER BY ts
      DEFINE 
        new_attempt AS (prev_rpm=0 AND rpm > 0)
        , low_rpm AS (rpm > 0 AND rpm < 750)
        , aborted AS (prev_rpm > 0 AND rpm = 0)
      PATTERN p AS (new_attempt low_rpm* aborted)
    ROWS MATCH FIRST EVENT
    )
  )
SELECT * FROM interesting_rows;
```
The clause has an OVER() sub clause, like any OLAP function, to specify along which column(s) to reset the series and which column to use as the ordering time value; then, you define a series of Boolean expressions and give those expressions event names. Finally, you specify the event series pattern you are searching for: (new_attempt low_rpm* aborted), meaning, like in a regular expression, one row of new attempt, zero to more of low rpm, and one of abort.

The functions visible in bold above depend on the MATCH() clause, and have the effect below:

<table>
<thead>
<tr>
<th>prev_rpm</th>
<th>test_drive_id</th>
<th>vehicleid</th>
<th>ts</th>
<th>rpm</th>
<th>airmass</th>
<th>event_name</th>
<th>pattern_id</th>
<th>match_id</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>27</td>
<td>3</td>
<td>00:05:10</td>
<td>300</td>
<td>5</td>
<td>new_attempt</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>300</td>
<td>27</td>
<td>3</td>
<td>00:05:20</td>
<td>400</td>
<td>6</td>
<td>low_rpm</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>400</td>
<td>27</td>
<td>3</td>
<td>00:05:50</td>
<td>430</td>
<td>7</td>
<td>low_rpm</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>430</td>
<td>27</td>
<td>3</td>
<td>00:06:00</td>
<td>340</td>
<td>4</td>
<td>low_rpm</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>340</td>
<td>27</td>
<td>3</td>
<td>00:06:12</td>
<td>40</td>
<td>1</td>
<td>low_rpm</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>40</td>
<td>27</td>
<td>3</td>
<td>00:06:23</td>
<td>0</td>
<td>0</td>
<td>aborted</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>0</td>
<td>29</td>
<td>3</td>
<td>00:06:17</td>
<td>300</td>
<td>5</td>
<td>new_attempt</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>300</td>
<td>29</td>
<td>3</td>
<td>00:06:24</td>
<td>400</td>
<td>6</td>
<td>low_rpm</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>400</td>
<td>29</td>
<td>3</td>
<td>00:06:37</td>
<td>430</td>
<td>7</td>
<td>low_rpm</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>430</td>
<td>29</td>
<td>3</td>
<td>00:06:58</td>
<td>340</td>
<td>4</td>
<td>low_rpm</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>340</td>
<td>29</td>
<td>3</td>
<td>00:07:12</td>
<td>40</td>
<td>1</td>
<td>low_rpm</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>40</td>
<td>29</td>
<td>3</td>
<td>00:07:23</td>
<td>0</td>
<td>0</td>
<td>aborted</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>0</td>
<td>29</td>
<td>3</td>
<td>00:07:51</td>
<td>300</td>
<td>5</td>
<td>new_attempt</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>300</td>
<td>29</td>
<td>3</td>
<td>00:08:01</td>
<td>400</td>
<td>6</td>
<td>low_rpm</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>400</td>
<td>29</td>
<td>3</td>
<td>00:08:17</td>
<td>430</td>
<td>7</td>
<td>low_rpm</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>430</td>
<td>29</td>
<td>3</td>
<td>00:08:25</td>
<td>340</td>
<td>4</td>
<td>low_rpm</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>340</td>
<td>29</td>
<td>3</td>
<td>00:08:43</td>
<td>40</td>
<td>1</td>
<td>low_rpm</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>40</td>
<td>29</td>
<td>3</td>
<td>00:08:57</td>
<td>0</td>
<td>0</td>
<td>aborted</td>
<td>2</td>
<td>6</td>
</tr>
</tbody>
</table>

Note that we only have a pattern ID value of 0 in test drive # 27, and pattern ID values of 0 and 1 in test drive # 29, as we experience the pattern twice in the second test drive.
APPLYING TIME SERIES ANALYSIS ON AIRCRAFT TRANSPONDER DATA

A broken stream of data

Using Automatic Dependent Surveillance Broadcast (ADS-B) data captured from transponders on commercial aircraft, we have found a perfect source of streaming data to demonstrate some of the data preparation, advanced analytics and machine learning capabilities of Vertica.

Capturing ICAO\(^1\) aircraft registration numbers, altitude, longitude, latitude and time stamp, as planes fly within range (200,000 KM\(^2\)) of our equipment located just 5 KM from Geneva airport, we are able to plot their tracks in near real-time. By persisting this data in a Vertica database, we can quickly build up a valuable source of historic flight data.

From this single stream of data (aka “messages”), identifying 10s, sometime 100s of aircraft in the sky at any moment in time, we observe messages arriving as often as every second for each aircraft. This can result in over 1,000,000 messages per day.

With messages arriving at varying time intervals, and no obvious indication from this data to distinguish different flights of the same aircraft, we will look at how Vertica’s Gap Filing & Interpolation and Sessionization can address these two problems.

The use case described below has been based on 4 months of captured data. This equates to ~130m messages from over 9,000 distinct aircraft. Although we will be executing queries against the full data set (no need to down sample when using Vertica!), for the purposes of simplifying this example, we will concentrate on just one aircraft – an 18 year old British Airways Airbus A319 with a ICAO registration number of “400801” (and call sign of “G-EUPA”). This aircraft was selected as it had been observed to both land/take-off from Geneva and fly over Geneva enroute to other destinations, providing us with ~33K messages during this period.

There are gaps to be filled

Looking at the messages received for this aircraft, it is clear to see the sporadic nature of the frequency the messages arrive. Sometimes every second, other times several seconds apart. Then after

\(^1\) https://en.wikipedia.org/wiki/Automatic_dependent_surveillance_%28%22ADS-B%22%29
https://www.icao.int/
21st July at 13:48:33, we next see this aircraft on 24th July at 22:23:18. Although this is the same physical aircraft, it clearly is not the same flight.

By using Vertica TIMESERIES function we can easily fill in these gaps (in the example below, to a time slice of 1 second), and to derive the altitude, latitude and longitude using a LINEAR scaling function.

However, just using this TIMESERIES function alone, we will also fill in all the gaps between the 21st July and 24th July, when as mentioned above, these are clearly two separate flights. To overcome this problem, we call upon Vertica Sessionization capability using a function called CONDITIONAL_TRUE_EVENT. This allows us to define a period (in this example 1 hour), which is used by Vertica to distinguish separate sessions (flights).

The result of combining the two functions, provides the following:

<table>
<thead>
<tr>
<th>X: x, y, z</th>
<th>Y: flight</th>
<th>Z: lat/long</th>
<th>GFI</th>
<th>Avro 55007</th>
<th>2017-07-21 13:30:01</th>
<th>16,100</th>
<th>47.4423</th>
<th>7.27107</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>400801</td>
<td>2017-07-21 13:30:02</td>
<td>16,100</td>
<td>47.44232</td>
<td>7.27107</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>400801</td>
<td>2017-07-21 13:30:03</td>
<td>16,100</td>
<td>47.44231</td>
<td>7.27106</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>400801</td>
<td>2017-07-21 13:30:04</td>
<td>16,100</td>
<td>47.44230</td>
<td>7.27105</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>400801</td>
<td>2017-07-21 13:30:05</td>
<td>16,100</td>
<td>47.44230</td>
<td>7.27105</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Plotting the tracks of this aircraft using the before (left) and after (right) data sets, not only can we see much clearer tracks due to the gaps being filled (GFI), but with the introduction of Sessionization, we are also able to distinguish different flights (sessions) this aircraft has flown – depicted here in colour.
Want to know more?

This use case forms part of an on-going project which has been used as a discussion topic at a series of Big Data and Machine Learning Meetups in London², Cambridge³ and Munich⁴.

The material presented at these Meetups is now also appearing as a series of blog postings on myVertica.com⁵.

Alternatively, please feel free to reach out directly to me at mark.whalley@microfocus.com.

⁵ https://my.vertica.com/blog/blog-post-series-using-vertica-track-commercial-aircraft-near-real-time/
THE VERTICA ANALYTICAL DATABASE: KEY FEATURES

1. **COLUMNAR STORAGE**
   Speeds query time by reading only necessary data.

2. **COMPRESSION**
   Reduces costly I/O & saves data storage space, leading to lower total cost of ownership.

3. **MPP SCALE-OUT**
   Provides high scalability on clusters with no name node or other single point of failure.

4. **DISTRIBUTED QUERY**
   Any node can initiate the queries & use other nodes for work.

5. **PROJECTIONS**
   Combine high availability with special optimizations for fast query performance.

VERTICA’S PORTFOLIO CONTAINS THE BROADEST RANGE OF DEPLOYMENT & CONSUMPTION MODELS

- **CLOUD**
- **ON-PREMISE**
- **APACHE® HADOOP**

ROBUST 3RD PARTY TECH SUPPLIER ECOSYSTEM, INCLUDING:

- **EASY OPEN SOURCE INTEGRATION**
- **BI/DATA VISUALIZATION**
- **DATA TRANSFORMATION**
- **ADVANCED ANALYTICS**

DATA MANAGEMENT AS A KEY ENABLER OF SMART SYSTEMS

Companies that wish to build Smart Systems business models based upon true, across-the-board digital automation, need to start with a data management & analytics solution that is purposely built to handle IoT data. The Vertica Analytical Database is well suited to be the key enabler that allows equipment manufacturers & IoT platform providers to increase the service scope & value delivered to their end-customers through system enhancement & optimization installed base data management, predictive analytics, and other increasingly complex use cases.
1. Streaming data API for Kafka: An API framework simplifies user data pipelines and allows for easy integration with Apache Kafka.

2. Integration with open source solutions: Easy integration with increasingly popular open source streaming data solutions like Apache Spark.

3. Management of semi-structured data: Manage hierarchical and messy semi-structured data sources by using Flex Tables, which do not require data schema creation and definition.

4. Geospatial data analytics: MPP architecture easily handles geospatial data from mobile sensors and conducts a wide variety of spatial analytics queries in fractions of a second.

5. Time-series analytics: Arranges data onto the same time scale and uses in-database analytical functions to interpolate missing values and get rid of outliers.

6. Built-in machine learning capabilities: In-database machine learning capabilities allow users to build models and train them on historical data.
A PLATFORM TO FOSTER MULTIDISCIPLINARY DATA SCIENCE COLLABORATIONS

Our abilities to collect, store and analyse scientific data have skyrocketed in the past decades, nevertheless at the same time, a divide between data scientists, domain experts and data providers has begun to emerge. Data scientists are progressively developing more powerful algorithms for data mining and analysis, while data providers are increasingly making data publicly available, and yet many, if not most, discoveries are based on specific data and/or algorithms that “are available from the authors upon request”.

In the strong belief that scientific progress would be considerably faster if reproduction and re-use of such data and algorithms were made easier, the Swiss Data Science Center is committed to provide an open framework for handling and tracking scientific data and algorithms, from raw data and first principle equations to final data products and visualisations, modular simulation models and benchmark evaluation algorithms.

The below poster demonstrates the vision of a high-scalable open but secure community-based platform for sharing, accessing, exploring, and analysing scientific data in easily reproducible workflows, augmented by automated provenance and impact tracking, knowledge graphs, fine-grained access right and digital right management, and a
variety of domain-specific software tools. For maximum interoperability, transparency and accessibility, notebook interfaces such as Apache Zeppelin and Jupyter are utilized wherever possible.

The preview beta version of this platform, named RENGA (連歌, a genre of Japanese linked-verse poetry in which two or more poets supply alternating sections of a poem linked by verbal and thematic associations), was launched in September.

In addition to facilitating multidisciplinary collaborations, the platform offers easy integration by leveraging proven, non-intrusive technologies as well as simplified compute and storage resource management.

https://datascience.ch/renga-platform/

The Swiss Data Science Center is a joint venture between EPFL and ETH Zurich. Its mission is to accelerate the adoption of data science and machine learning techniques within academic disciplines of the ETH Domain, the Swiss academic community at large, and the industrial sector. In particular, it addresses the gap between those who create data, those who develop data analytics and systems, and those who could potentially extract value from it. The center is composed of a large multidisciplinary team of data and computer scientists, and experts in select domains, with offices in Lausanne and Zurich.

www.datascience.ch
This article aims to give an overview of analytics as a driver of performance in banking. We will focus here on activities related to financial intermediation, i.e. banks as intermediaries between depositors and borrowers.

Let us take the example of the Swiss market. As shown in Table 1, as of today, half of the global balance sheet of the banks in Switzerland is still dedicated to Intermediation (items highlighted in grey), where Mortgage comprise over 30% of the aggregate balance sheet total. Table 1 aggregates data for total domestic and foreign Banks in Switzerland:\(^1\)

In this context, analytics play a fundamental role to improve banks Profitability (I), Stability (II) and Attractivity (III).

I – PROFITABILITY

Currently, at least two phenomena harm the profitability of the banks on financial intermediation. The first is linked to the low rates that have been going on for several years now\(^2\). The second is linked to the entry of new market players (shadow banks, crowdlenders, GAFA, Telco... the list goes on). In this unfavorable economic context, banks can leverage on cost reduction through digitalization, robotization and analytics to improve their long-term and short-term profitability.

First, and that is a trend that can be observed since several years in the banking industry but also in other industries, the costs (of labor, real estate...) decrease with the development of automated and digital services. In the banking industry, each of us can observe the development of e-banking and digital payments, but also credit. Indeed, credit is a critical activity for Swiss banks. As shown in Table 1, loans weight for half of their total assets (CHF 1.541 billion). According to the size of the client, the amount of the loan, and the complexity of the operation, credit process can be industrialized with a direct impact on operational costs. It also allows to improve operational efficiency, especially to higher the volume, to equal risk, through an acceleration of the credit approval process and even, on basic loans, access to credit 24/7.

---

\(^1\) Including: Big banks, Cantonal banks, Raiffeisen, Private bankers, Foreign-controlled banks, branch of foreign banks, Regional & Savings banks, Stock Exchange banks and other banking institutions.

\(^2\) At the time when we write this article (sept 2017), the Swiss rates are still negative.

---

Table 1 - Balance Sheet - Total domestic and foreign Banks in Switzerland (CHF billion)

<table>
<thead>
<tr>
<th>ASSETS</th>
<th>LIABILITIES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total assets</td>
<td>CHF 3,104.82 billion</td>
</tr>
<tr>
<td>Total liabilities</td>
<td>CHF 3,104.82 billion</td>
</tr>
</tbody>
</table>

Source: Swiss National Bank (SNB) – Bankstat Switzerland 2018 – table of annual banking statistics
Second, banks leverage on analytics to lower the cost of risk linked to the uncertainty of their activity. For credit, this risk is materialized by financial losses if loans turn to be non-performing. This cost of risk can weigh heavily on their profitability, so the robustness of models and their ability to counteract change is crucial. Hence, the improvement of tools along with the improvement of the models based on machine learning, the introduction of real time data and unstructured data can help improve analytics at each step of the credit process, and thus reduce uncertainty and cost of risk: at launch to calibrate acceptance (obligor rating, scoring...), and more generally to define acceptance rules (yes/no, interest rate, covenants...), during the life of the credit to monitor the non-performing loans (including: expected losses, provisions and recovery process requirements), but also at a portfolio level, where analytics are performed to identify global risks such as concentration (on a country, on a sector, ...) and determine limits and watch lists.

II – STABILITY

Financial intermediation is the primary role of banks. They transform the maturity and the liquidity of the deposit to turn it into loans. Hence, Asset Liability Management (ALM) helps define strategies to optimize profitability while ensuring long-term stability, taking into account markets evolution and customers behaviors. Banks perform supervision to ensure their own stability, but also to ensure the global system stability as some banks are said to be systemic\(^3\), meaning that their failure can affect the entire system. Here the definition of financial stability by the European Central bank: "Financial stability is a state whereby the build-up of systemic risk is prevented. Systemic risk can best be described as the risk that the provision of necessary financial products and services by the financial system will be impaired to a point where economic growth and welfare may be materially affected."\(^6\).

The issue is such that stability supervision is strongly regulated, especially since the 2008 crisis. Compliance not only means compliance with ratios, reporting and stress tests, but also compliance of the quantitative models (for example Basel regulation about internal vs standard models) and of the internal processes of risk supervision themselves (for example BCBS239 about risk data aggregation, reporting and data governance).

To deal with the strengthening of regulatory requirements, banks can rely on the development of tools in the field of RegTech. Even more, the regulatory pressure has increased so much, requiring crossing of data in all directions, that some banks wonder if they should stop aggregating data as they did before (based on “classical” financial framework) and switch to a model where they collect and store granular data to be free to aggregate them “on demand” according the evolution of regulation\(^5\). This change in methodology could benefit from the improvement in Big Data storage technologies and analytics.

III – ATTRACTIVITY

As shown in table 1, customer deposits represent more than the half of the Swiss banks resources (CHF 1,771 billion). Thus, to maintain an important level of stable deposits and ensure the stability of their balance sheet, banks need to be attractive in a highly disruptive and competitive environment. Based on customer and transaction data, banks carry out marketing analysis and customer segmentation to propose innovating and convenient products and services, and maximize equipment rates.

To do this, they rely on a large amount of customer information that they collect since a very long time, in comparison with other industries. Due to the specificities of banking activities (regulations requirements about KYC (Know Your Clients),

\(^3\) Banks can be systemic at a global –G-SIB– or at a domestic –D-SIB– level
\(^6\) Source: ECB website.
\(^5\) See Revue Banque N°801 novembre 2016, dossier Reporting Réglementaire.
Are you tired of all these books about Big Data? They are all very similar. They claim that the field is brand new, the time is now, the data is huge and the maths are complex. Weapons of Math Destruction (WMD) is a unique book and a fresh reading, to say the least. Cathy O’Neil writes about Big Data from a different angle, and it will be a nice addition to your Data Science bookshelf.

Cathy discusses how Big Data, or any data-driven application, can create inequality issues and threaten democracy. This sounded provocative and exaggerated to me while starting the book. Now my opinion is enlightened by the examples provided throughout the book. A Weapon of Math Destruction – as defined by Cathy – is an algorithm that, due its opacity and scale, causes damages to people lives.

The concept of WMD is very interesting, mainly because we usually have in mind good consequences of Big Data: find frauds, fight crime, improve health research or transform a company for the better. For this to happen, we collect huge amount of data. We also run black-box models that take important decisions. What about ethics and privacy? WMD proposes to address these topics through real cases.

Cathy shares plenty of examples from her own experience. From teaching to finance and call centers to job interviews, examples are heterogeneous and comprehensive. At first, one may think that a computer-based decision is more objective than a Human one. Cathy is arguing that the Human bias is hidden in the technology. To add to this, models are readable – if at all – by an elite only.

At that point, a legitimate question comes to mind: why is the author still in the field of Data Science? The answer is that you can avoid WMD. Cathy is fighting for understandable models and showing ways to avoid creating WMD. Whether you want to reduce challenges caused by such WMD tools or simply open your mind with a fresh reading, just go for it.

CONCLUSION

Bank is structurally a data-driven business, we can even say that it is a mature industry in term of analytics. The emergence of more data – data everywhere –, more agile tools that rely on cheaper technologies provides banks opportunities to improve their profitability and stability along with the attractiveness of their products and services.
SUPPORT THE SWISS ASSOCIATION FOR ANALYTICS!

If you like our association and would like to contribute, you can write an article for our magazine, be a speaker at one of our events or start a discussion on our LinkedIn group:

LinkedIn: www.linkedin.com/groups/4586163
E-mail: info@swiss-analytics.com
Meetup: www.meetup.com/swiss-analytics

As a non-profit organization, we would also welcome any donation. Your participation will help us cover regular costs (web hosting, meetup subscription, etc.)

Our bank account is:

Name: Swiss Association for Analytics
Address: Chemin des Fontannins 12, 1066 Epalinges
Bank: PostFinance
Account: 85-122140-9
IBAN: CH65 0900 0000 8512 2140 9
BIC: POFICHBEXXX

Thanks for your support!

IMPRESSUM

The Swiss Analytics Magazine is proposed by the Swiss Association for Analytics (www.swiss-analytics.com). The magazine is printed by Gasser Media (www.gasser-media.ch). Articles are written by experts in analytics. The views expressed in this magazine do not necessarily represent the views of the Swiss Association for Analytics. It is not allowed to reproduce, duplicate, copy, sell, resell or exploit any portion of the magazine without agreement of the Swiss Association for Analytics (info@swiss-analytics.com).

Printed in January 2018, Le Locle, Switzerland
LE MEILLEUR DE VOS DONNÉES !
COLLECTER, ORGANISER, DIFFUSER.

ÉDITION
GESTION DE PROJET
DEPUIS 2001

IMPRESSION
PRODUCTION À LA DEMANDE
DEPUIS 1947

PUBLICATION
WEB, APP, CROSS-MEDIA
DEPUIS 2004

ADRESSE
Gasser Media SA
Jambe-Ducommun 6a
CH-2400 Le Locle

CONTACT
Tél. +41 32 933 00 33
info@gasser-media.ch
www.gasser-media.ch

Pour suivre nos actualités :
www.gasser-media.ch/blog