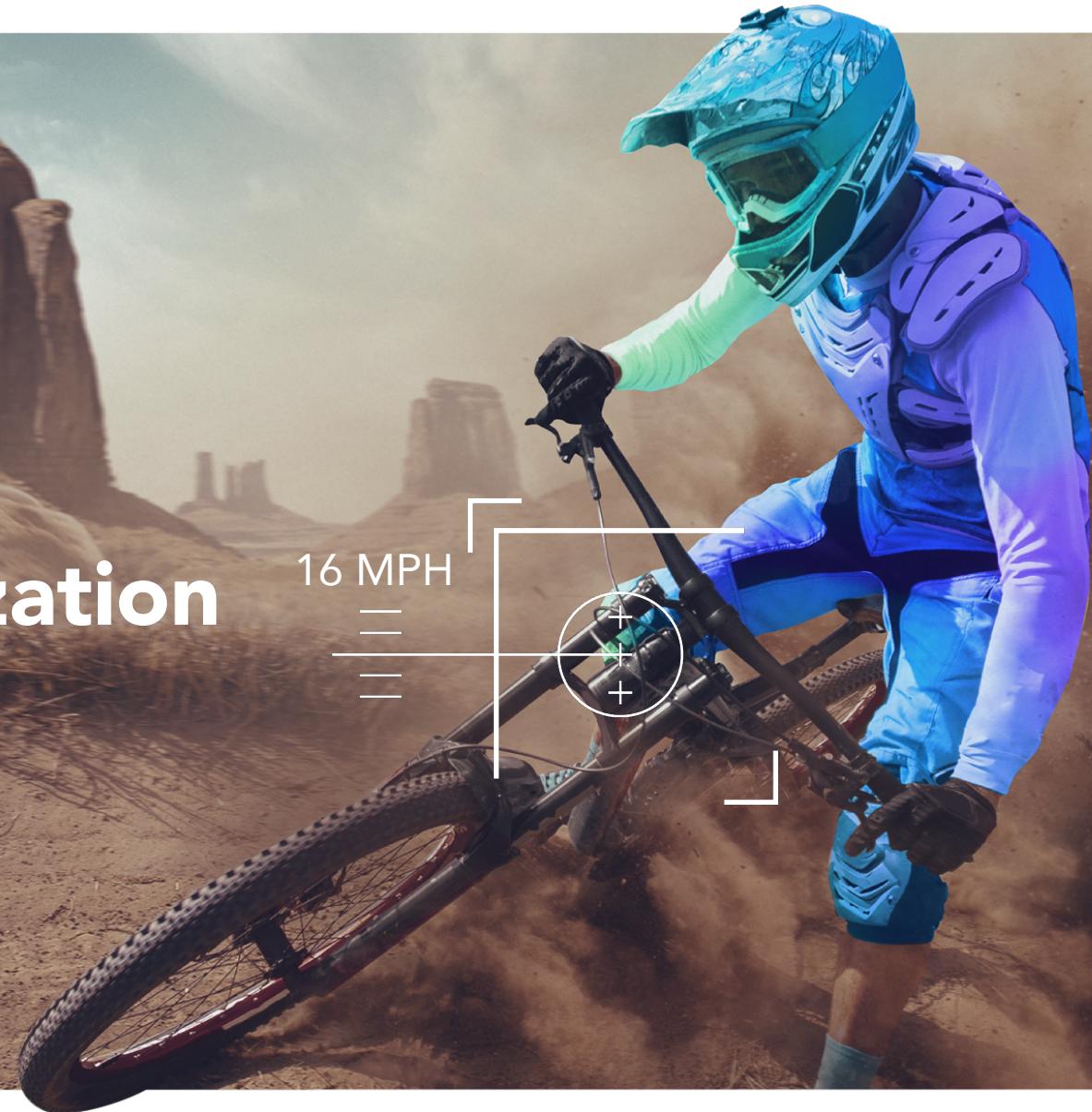




STARSCHEMA

We help your organization
become data driven



Five ways to leverage AI in Tableau

Tamás Földi
CTO, Starschema

Chris Csefalvay
VP Special Projects, Starschema

I About us

Your presenters

About Tamás



Tamás Földi

CTO and Partner, Starschema

- Co-Founder and Partner at Starschema
- Tableau Zen Master for the fourth consecutive year

About Kristóf



Dr Kristóf Cséfalvay

VP Special Projects, Starschema

- Originally trained in law, then in clinical epidemiology
- 10-year track record of data science jobs, including at companies like Volkswagen AG, RB plc and various roles in public service
- Special interest in neural networks/deep learning, geospatial analytics and data exploration
- Joined Starschema as Principal Data Scientist, currently serving as VP Special Projects

I AutoML in Tableau

Explain Data in Tableau

I NLP in Tableau

Analysing customer reviews for the cause of a rating

About NLP



- Natural Language Processing (NLP) is the data science discipline that pertains to extracting information from, and learning on, unstructured or semi-structured textual data:
 - computer-generated text: logs, status messages,...
 - VOC (Voice of the Customer): e-mails, call transcripts, chatbot interactions, customer reviews...
 - social media, brand surveillance, topic surveillance...
 - internal and external knowledge bases, documents, reports, legal documentation (contracts, deeds &c.)...
- Based on research in computational linguistics since the 1970s, it has been enormously successful in a range of applications – to mention only a few examples:
 - determining sentiment from a large volume of data, e.g. thousands of product reviews or tweets,
 - categorising institutional knowledge by the most pertinent key concepts,
 - analysing error tickets to prioritise support skillset,
 - understanding ratings from data sets that couple a rating with a review.

Example data set: women's clothing reviews



- 23,486 reviews of women's clothing from an unnamed e-commerce distributor, anonymised
- Variables:
 - Review text
 - Rating (numeric, 1–5)
 - Review weighter (number of users who have found a review to be useful)
- Approach: tf/idf association
 - Isolates terms that are relatively frequent within one review when compared to all reviews
 - In other words: what terms are particularly salient to
 - each review,
 - good reviews (rating 4 or 5),
 - bad reviews (rating 1 or 2)
 - From this, we can create a picture of what drives customers' interest

I **Seasonality detection in Tableau**

Detect seasonal components and trends in weather data

About seasonality detection



- Some processes over time exhibit a combination of a periodic effect (seasonality) and an aperiodic effect (trend), combined with a noise variable to account for everything else.
- Consider air traffic passenger counts:
 - Overall, as air travel has become more affordable, more people travel by air today than 10 years ago (aperiodic effect/trend).
 - Every year, people generally travel more by air during Christmas and Summer school holidays than at other times of the year (periodic effect/seasonality, since it is reflected in every year)
 - Beyond this, there are variations that aren't acutely accounted for by either seasonal or trend effects – a particularly bad air traffic disaster, for instance, may result in reduced bookings (accounted for by the noise variable)
- Seasonality detection allows
 - a better understanding of the cyclic processes that drive demand effects,
 - while also revealing more about trends while controlling for seasonal effects.

Anomaly detection in Tableau

Spot anomalies in engineering telemetry

About anomaly detection



- Anomaly detection identifies data points that fall outside a 'normal' pattern:
 - abnormal combinations: people who buy bleach and duct tape (and nothing else)
 - time series anomalies; sudden spikes or outliers
- Many of these can be easily captured visually, but machine learning can be useful in identifying and quantifying them, in real time – allowing immediate response:
 - fraud prevention: acting on evidence of anomalous transactions by blocking them (e.g. credit card lockdowns after a sudden sequence of cash withdrawals)
 - maintaining quality of service: responding to increased demand due to extraneous factors by bringing in additional resources
 - predictive maintenance and fault detection: using anomalies to detect prodromic signs of impending breakdown
 - pattern break recognition: using an established pattern, such as average key press duration and average time between key presses within a word, as an 'identity' and alerting when this pattern is deviated from, a technique widely used to detect fraud on sensitive systems

Example data set: NAB Machine Temperature



- 22,695 data points consisting of measurements from a temperature within a machine, with a temporal resolution of 5 minutes. The accumulation of an initial error and increasing degradation led to catastrophic failure of the machine at the end.
- The data set is a good and very realistic stand-in for a single dimension of real-world predictive maintenance data we encounter on a regular basis – especially because it contains not one catastrophic anomaly but a sequence of a very overt and a much less overt anomaly culminating in catastrophic breakdown.
- Timely detection of such anomalies can
 - save repair costs,
 - reduce outages and disruptions to supply chain processes,
 - prevent catastrophic breakdowns, and
 - reduce the potential risk of personal injury or loss of property associated with serious defects.

Natural Language Generation in Tableau

Actionable intelligence in clear words

About Natural Language Generation



- Natural Language Generation (NLG) refers to the synthesis of natural language responses to data.
- NLG makes data intelligible by summarising it in a clear, cogent natural language output. While visualisations show what the data is, NLG is about automatically creating a data narrative based on the saliency of information. This saves the lengthy process of manually writing reports to stakeholders who rely on quick textual summaries.
- In the following demo, a finance dashboard will be accompanied by an NLG output using a solution developed by our partner, Arria. The NLG output will put the often complex visuals into plain text, while also tied closely to the data and being responsive to selections on the main dashboard.



STARSCHEMA

We help your organization
become data driven

