



Was kann Machine Learning eigentlich für den Alltag des SQL Server DBAs tun?


Sascha Lorenz

PSG Projekt Service GmbH | Hamburg | Karlsruhe
lorenz@psg.de





Agenda

- Journey of „Self Disruption“ 😊
- Machine Learning 101- Unsupervised als auch Supervised Learning
- Einsatz für die fortgeschrittene Überwachung & Optimierung von Datenbank Systemen (u.a. Microsoft SQL Server)
- Python  jupyter
- Knowledge / Risk / Change / Process Management 😊



<https://github.com/psglorenz/Sessions>



meetup

[Neue Gruppe gründen](#)

[Entdecken](#)


[Nachrichten](#)

[Updates](#)



Microsoft Data Platform Community Hamburg - More than SQL

 Hamburg, Deutschland

 548 Mitglieder · Öffentliche Gruppe

 Organisiert von **Sascha Lorenz** und 1 weiterer

Teilen:   

[Über uns](#)

[Events](#)

[Mitglieder](#)

[Fotos](#)

[Diskussionen](#)

[Mehr](#)

[Gruppe verwalten](#)

[Event erstellen](#)

Worum es bei uns geht

Dies ist die definitive Gruppe für alle die mit dem Microsoft SQL Server und der Microsoft Data Platform - Azure - PowerBI etc. zu tun haben! Wir sind ein...

Organisatoren



Sascha Lorenz und 1 weiterer

[Nachricht](#)

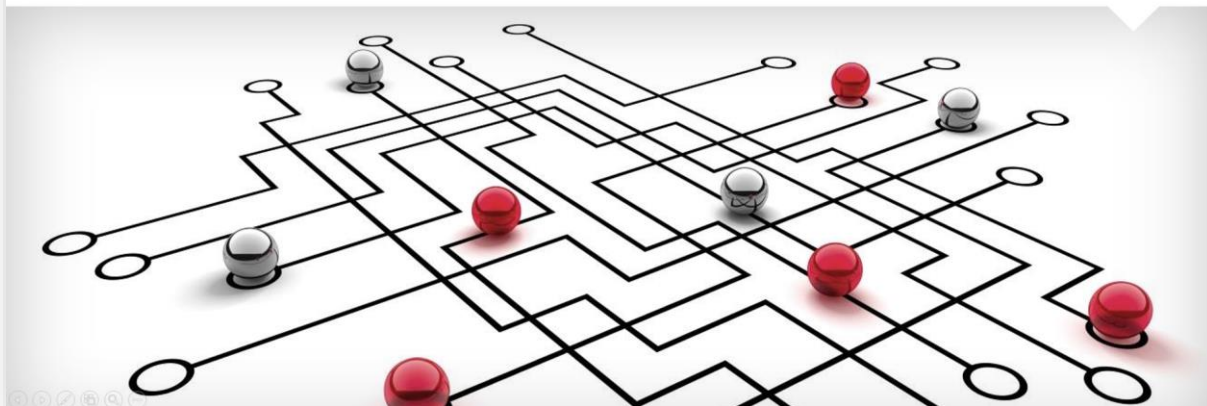


Unsere Motivation...

Complex Environments

Data in database systems are always part of a bigger picture. These data are often **distributed** over many databases and servers, in addition, they are not always adequately **fast** and **reliable**.

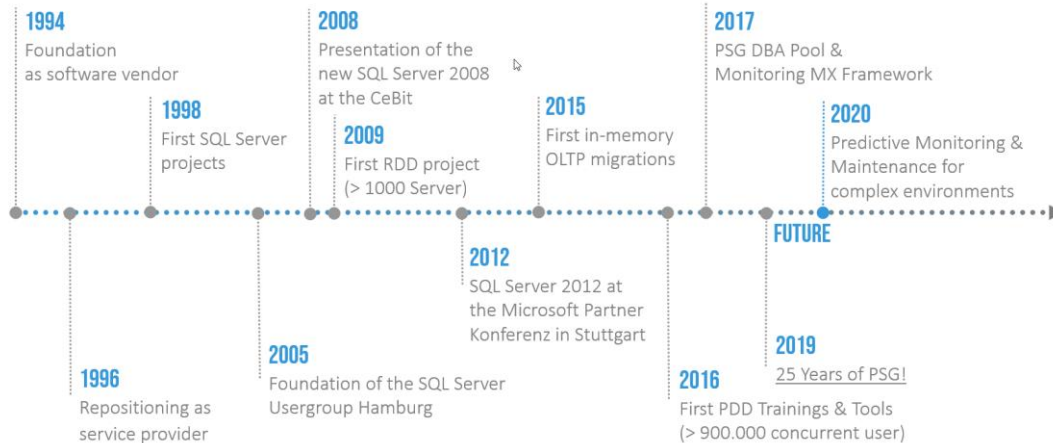
Such systems can become very complex over time!





Unsere Motivation...

Milestones





Unsere Motivation...

EXAMPLES

How do our customers use the capabilities of PSG to manage their (complex) environments?

PERFORMANCE

- Analyse and solve performance issues
- Plan and implement monitoring solutions (PSG PDD & PE)

COMPLEX DATA SILO

- Analyse silo dependencies
- Consolidate as well as decentralize structures (Orchestration RDD)

VENDOR AUDIT

- SQL Server best practices usage audit
- Communication with vendor's dev team

RELIABILITY

- Plan and setup HA features
- Create and implement operational concepts

MIGRATION

- From 20xx to newer versions & cloud
- From Cluster to AVGs (live)

KNOWLEDGE

- Plan and migrate to in-memory OLTP
- Design and create repositories to manage environments (RDD)





Was ist unsere Motivation?

- Ein „SQL“ Datenbank Server ist wohl eine der Komponenten mit dem höchsten operativen Risiko...
- Es gibt zwei klassische Szenarien:
 - Are we in trouble...?
 - We are under fire...!
- „Wir sind ja dabei, aber wo stehen wir gerade?“
- Wie sollten sich DBAs und Entwickler „organisieren“?
- Wie und worauf sollten im Team Prioritäten gesetzt werden?
- Was sollte wann wie im Team kommuniziert werden?



Eine unserer Antworten

- Ein Risikomanagement um den Query Optimizer & die Entwickler herum...
- **PSG PDD** – Performance Driven Development
 - Ein (technisches) Kommunikationsmodell, um die Entwicklung und den Betrieb von Datenbank Systemen im Team zu optimieren
 - Performance & Skalierbarkeit... (Advanced „State“ Management)
 - Dokumentation von Entscheidungen
 - Gelebtes DevOps



Unsere Motivation...

PSG METHODS & ASSOCIATED TRAININGS

Overview of our approaches

PSG PTS

PERFORMANCE,
TROUBLESHOOTING &
SUPPORT

The focus of this method is on instant analysis and problem solving

This method includes best practices and proven methods to analyze performance and stability challenges.

Target audience are administrators and one-man project teams

PSG PDD

PERFORMANCE
DRIVEN
DEVELOPMENT

The focus of this approach is on the performance aspect of the development or redesign of complex systems based on SQL Server.

The development of sufficiently fast software with SQL Server is seen in this approach as a series of risks.

This method provides actions which are taken by the team to mitigate these risks.

Target audience are developers, team leads and project managers.

PSG RDD

REPOSITORY
DRIVEN
DESIGN

The focus of this architectural and development approach is to mitigate the risks of getting lost in complexity.

The emphasis is on designing repositories to reduce complexity in large environments.

Such repositories enable the possibility of managing systems in more efficient ways.

The target group are companies with tough challenges regarding the number of servers, databases, processes and/or users.



Unsere Motivation...

DBA as a Service (DBA Pool)

Advantages

NO STAFF SHORTFALL

A team of experts takes care of your SQL Servers and processes.

PERIODICAL ASSESSMENTS

The DBA team regularly evaluates the possible risks of stability, performance and scalability.

COST TRANSPARENCY

Additional services are performed only after approval.

THE SQL SERVER COMPANY

PSG accompanies you during all imaginable SQL Server projects.

COMMUNICATION AS EQUALS

Our team communicates on an equal footing with your vendors and developers.

USAGE OF PSG MX

Your SQL Servers can be analyzed on demand with the monitoring framework PSG MX (xService & xVault).

Our valuable subject matter expert is ready to be involved in daily tasks as well as in projects.





Unsere Motivation...

PSG MX AIOps as a Service

Overview

FOCUS ON PROCESSES

PSG MX concentrates on the business processes which cause SQL Server workload. Behind the scenes, MX uses the PSG RDD method to group technical details of the workload to reduce the complexity of a solution.

RISK MANAGEMENT


MX focuses on the risks identified as significant and tracks whether there may be a danger to the work of the solution. In doing so, the emphasis is placed on understandable to all involved parties evaluations.

FLEXIBLE FRAMEWORK

Since the MX monitoring framework is a PSG in-house development, it is ensured that a maximum degree of toolset flexibility can be achieved for different requirements.

PAY PER DETAIL

PSG MX offers different levels of insights. Firstly, it provides risk management based information for managing a solution. Secondly, it can automatically deliver feedback about the quality of the solution to developers.





We are hiring!

Arbeiten Sie mit uns an unserem
DBA Monitoring Framework PSG MX

SQL Server | T-SQL | PowerShell | Python | TensorFlow | Keras | scikit-learn | ML.NET



Unsere Sicht auf den SQL Server

Dashboards...

Macro Level

(CPU, Batch Requests/s, Transactions, Storage Latency, Waits etc.)

Workload

„Queries...“

Top 42
Expensive
Queries..

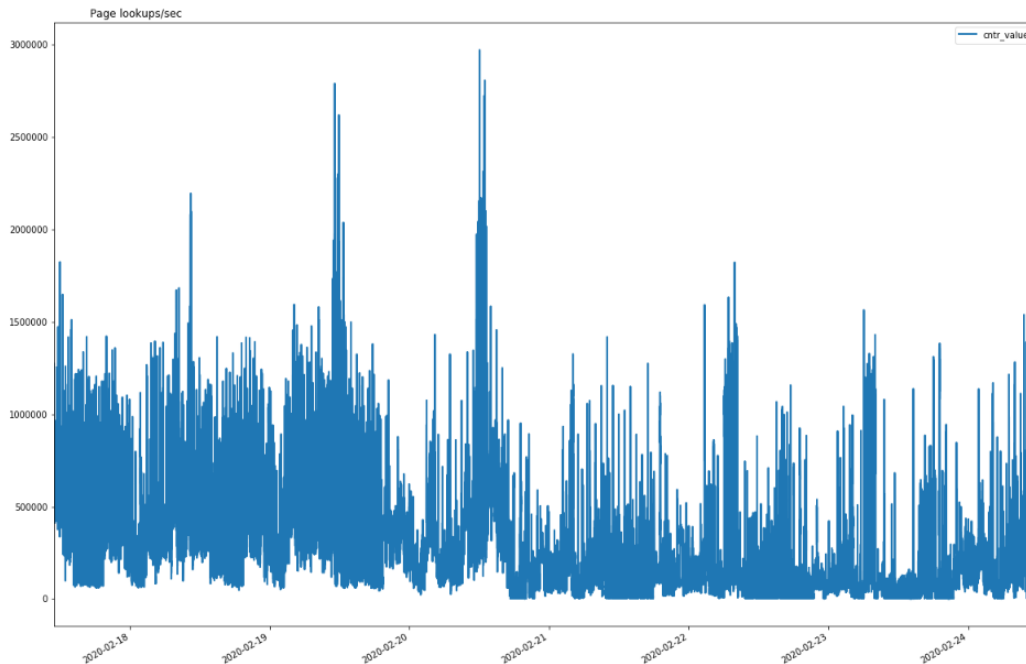
OMG!

Micro Level

Index „Usage“, Index Latches/Locks, Stats Updates, Rows Stats, Toxic Columns etc.)

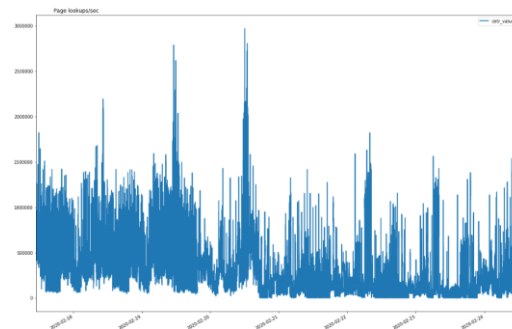
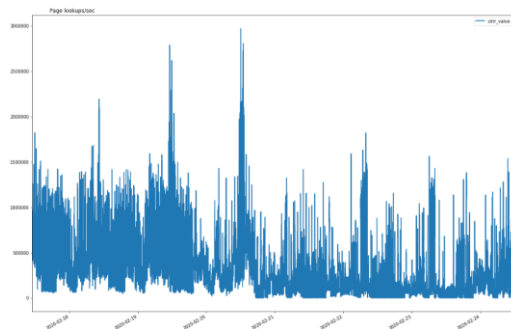
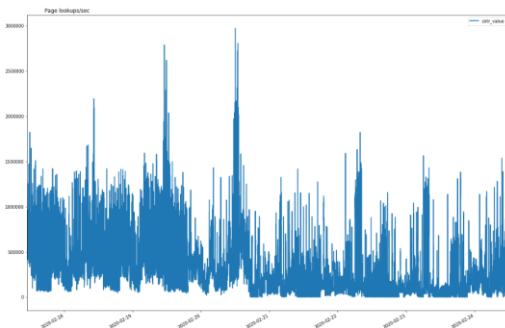
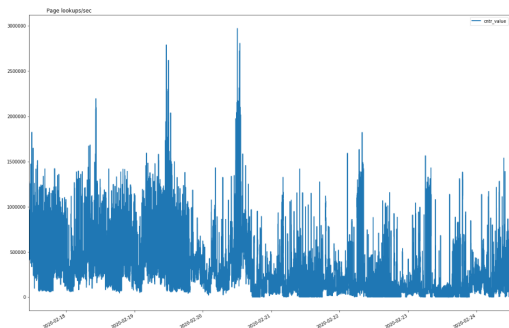


Monitoring...



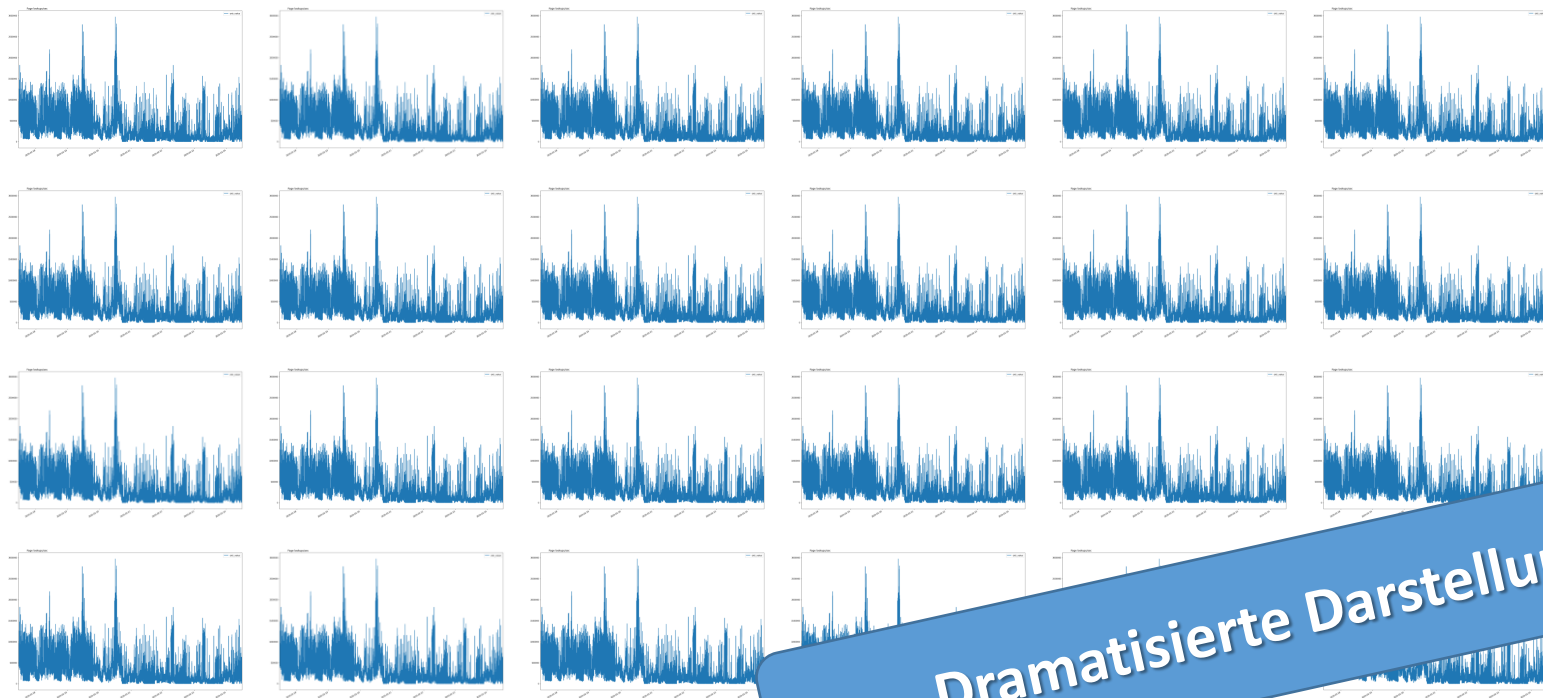


Monitoring...





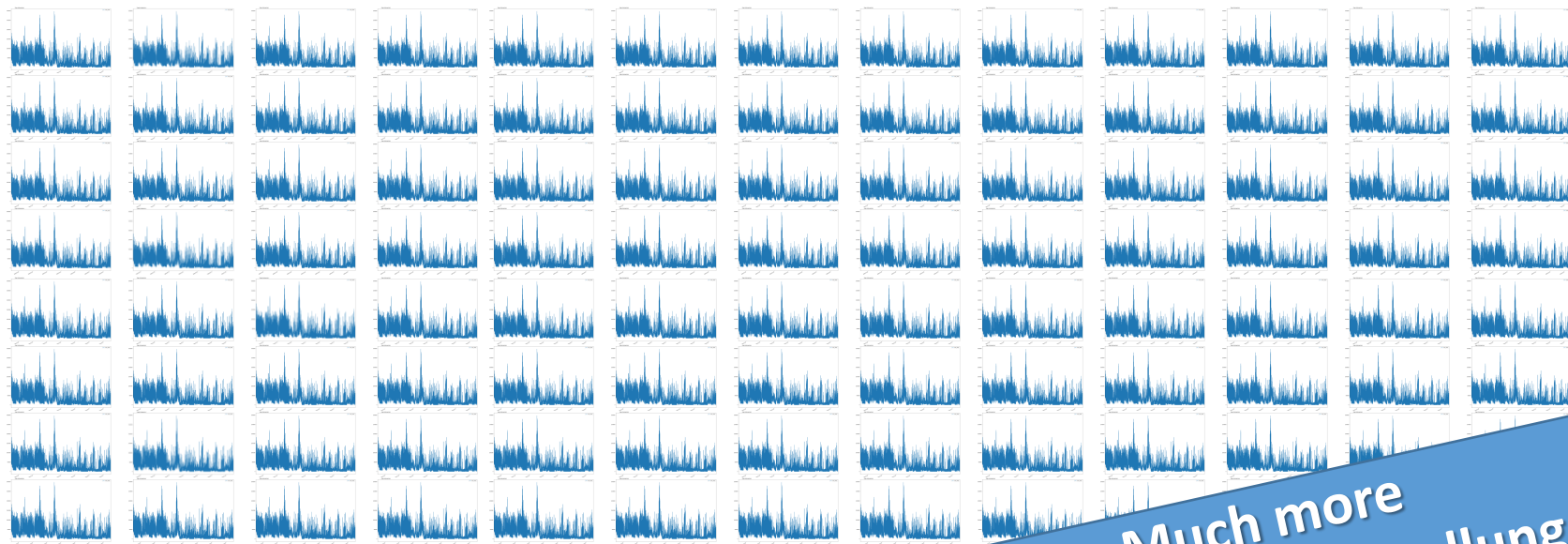
Monitoring...



Dramatisierte Darstellung...



Monitoring...



Much more
dramatisierte Darstellung...



Viele Daten...

- Schon eine einzelne SQL Server Instance kann recht schnell zu einer **Big-Data Herausforderung** werden!
- Wir sammeln zum Teil (primär in Health Checks/Audits) mehr als 50.000 Datenpunkte pro Datenbank/Min.
- Der Verzicht auf eine Visualisierung als der primäre Ansatz für eine Analyse befreit und macht kreativ!



Wie gehen wir das Thema generell an?

Scores

Tags



Scores

- Meist einfache Mathematik, auch mal Statistik
- Bilden von Verhältnissen zwischen diversen Metriken
- Beispiele:
 - SQL Prozess Auslastung zu Batch Requests/s
 - Verhältnis von Scans zu Updates auf einem Index
 - Rows im Verhältnis zu Logical Reads (mal in Frage stellen)
 - Und Viele mehr...



Tags

- Tags helfen uns die Server „multidimensional“ zu sehen
- Wir heften an quasi alles einen oder mehrere Tags
 - Server, Databases, Tables, Indexes, Queries, Plans, Columns etc.
- Tags können das Ergebnis eines Scores aber auch eines „simplen“ Alerts (z. B. Lock Escalation) sein
- Tags können aber auch durch die Klassifizierung durch ein ML Modell (Un- / Supervised) vorgenommen werden
- Beispiele für Tags:
 - (B)Lock hot spot, too many „scans“, toxic column etc.



Wirkungsketten

- Das System kennt auch diverse feste Wirkungsketten
 - Wenn das, dann das... und doch das?
 - Rule based AI 😊

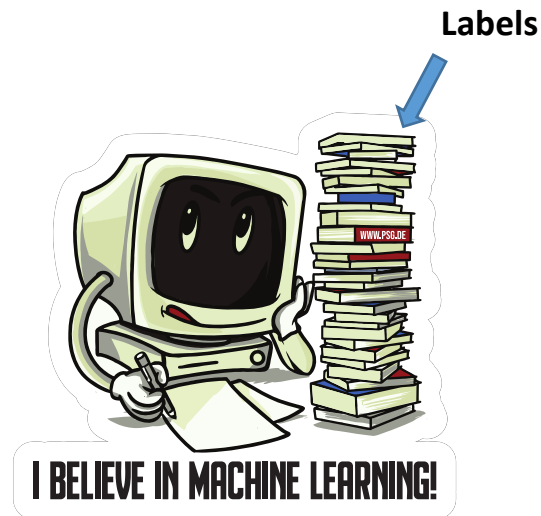


Machine Learning 101

- Generell wird unterschieden zwischen:

Supervised Learning
(wir haben „Labels“)

Unsupervised Learning
(wir haben keine Labels...)





Machine Learning 101

- Die „Welt“ durch Gleichungssysteme beschreiben, damit wir Zustände (Labels) vorhersagen können
- Dazu haben wir u.a. Daten als Input zum Lernen
- Diesen Input nennen wir ab jetzt Features!

$$\gamma = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_k X_k + \varepsilon$$

Labels

Feature

Feature

Feature

Gewicht

Gewicht

Gewicht

Gewicht

Fehler !



Mean Squared Error

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2$$



Arten des Machine Learnings

- Supervised Learning
 - Regression
 - Classification
 - Binary
 - 1-1
 - 1-n
- Unsupervised Learning (learning by observation)
 - Clustering
 - Dimensionality reduction (PCA)
- Reinforcement Learning
 - Q-Learning
- AutoML...



Wann sollte ich nun Machine Learning nutzen?

- Immer wenn es darum geht etwas „vorherzusagen“!
- Okay, okay... :)
- Neuer Versuch:
- Bei allem was mit klassischer Mathematik nicht umsetzbar ist.



Vorhersagen (Predictions)

- Anomaly Detection?
- Automatisierung von Entscheidungen!
- Gruppierung von Objekten
 - Indexe, Queries, Plans
- Korrelation & Muster Erkennung
 - Transactions, new/updated rows (stats), backup size, toxic columns
- Filterung des Alert Streams
- Feature Generierung!



PSG MX AIOps als Middleware



PostgreSQL



PSG MX

xVault

PSG DX

Nagios®

ZABBIX

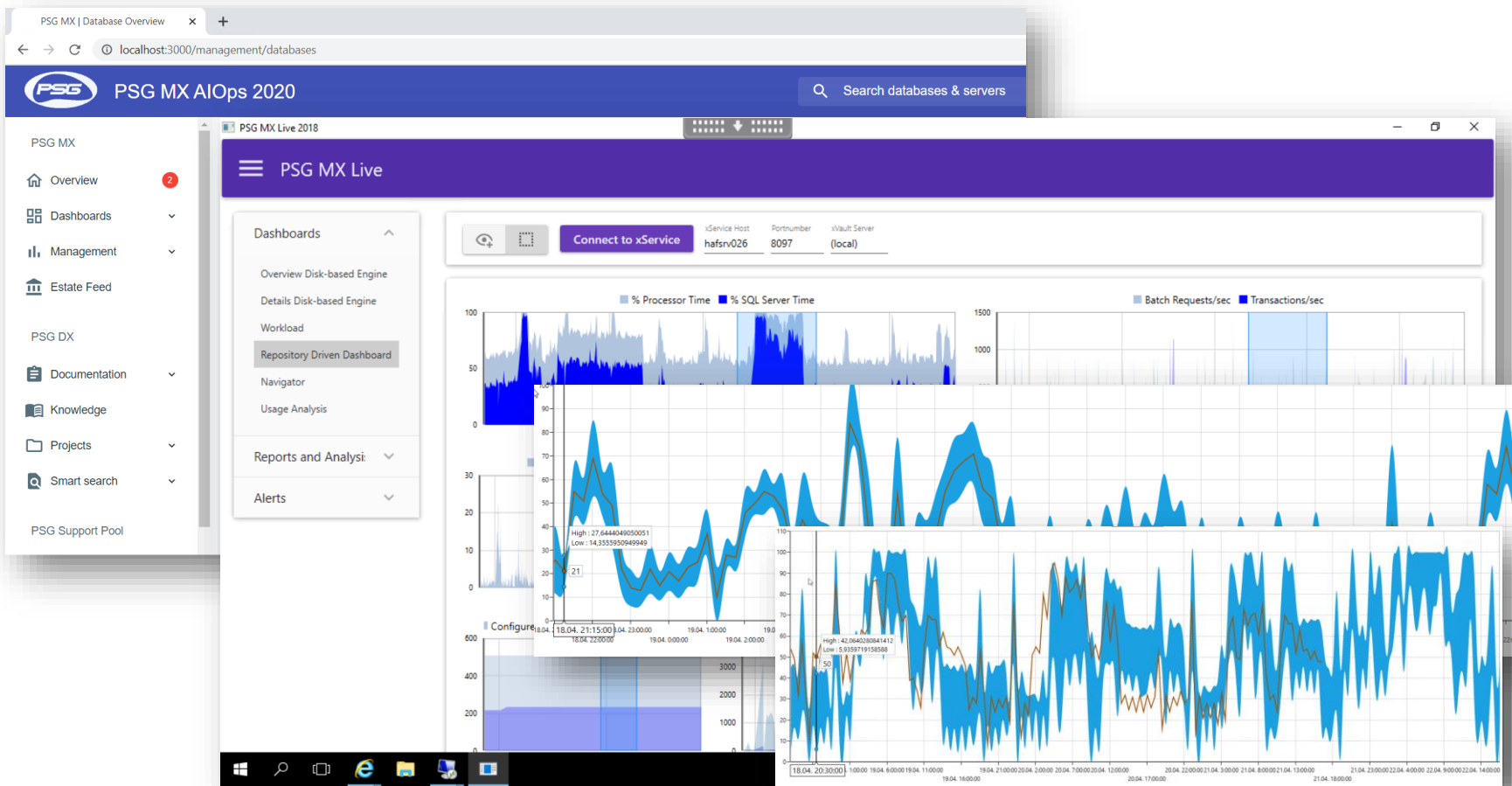
ICINGA





SQL Server Konferenz 2020

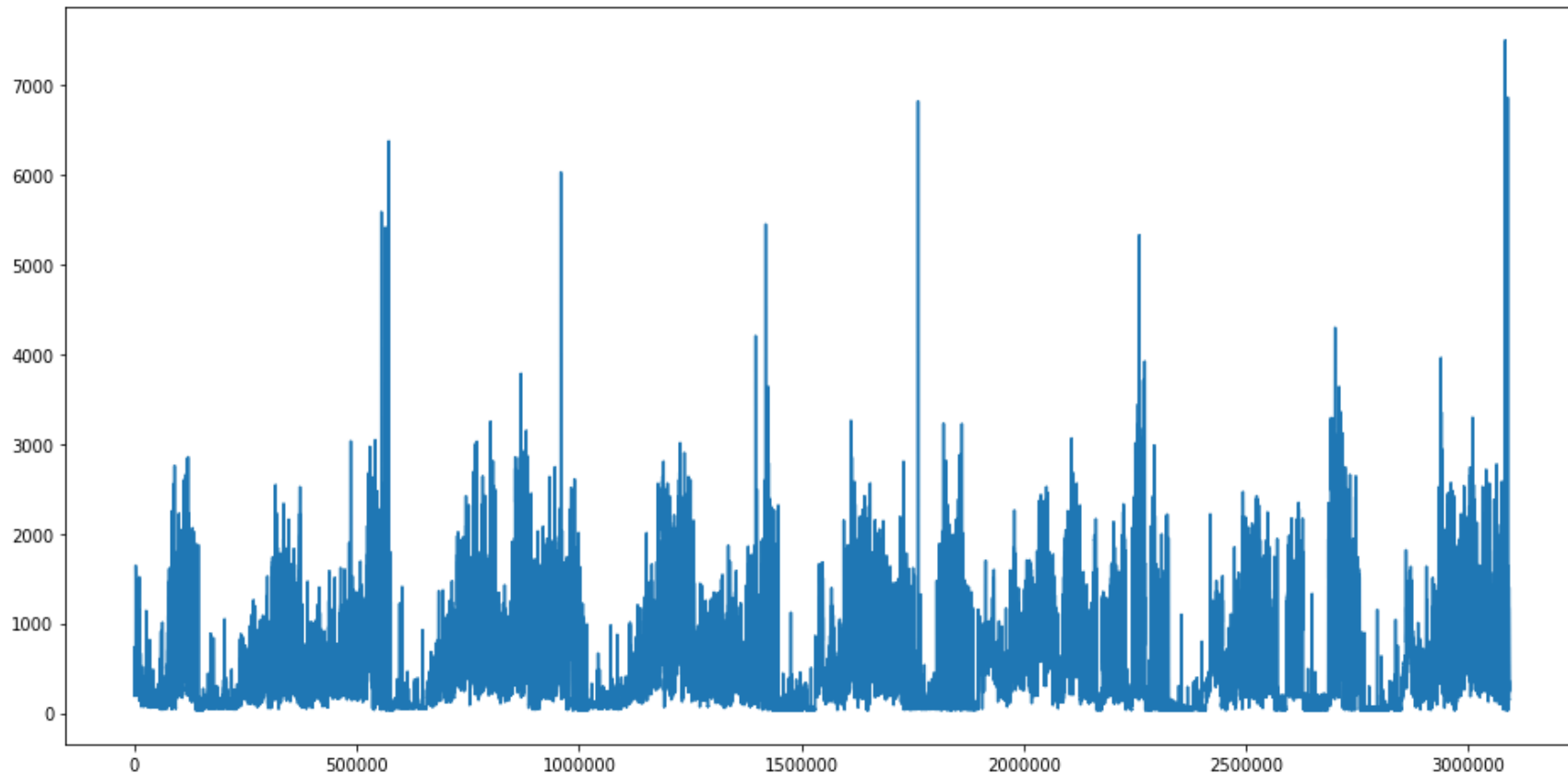
powered by PASS Deutschland e.V.

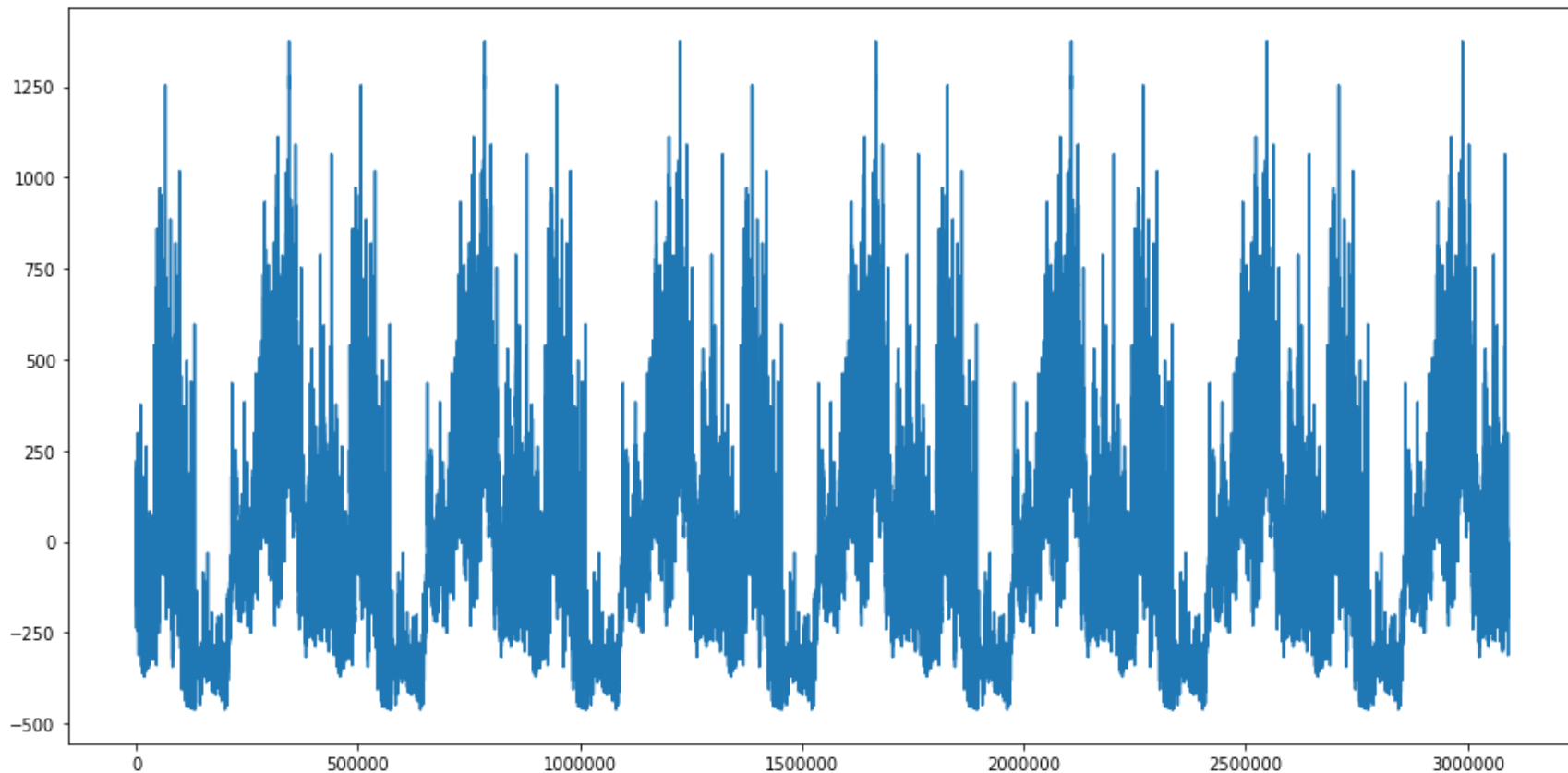


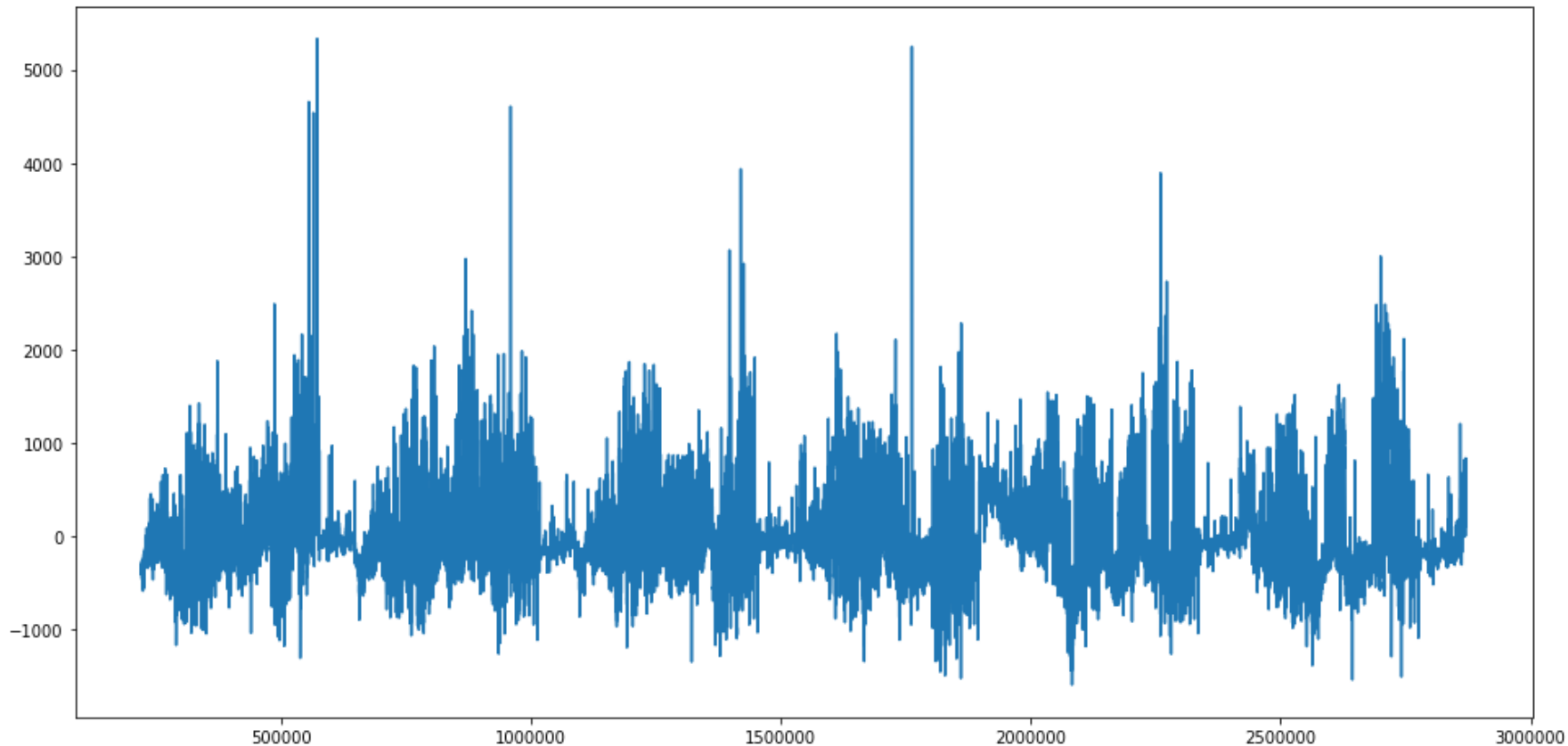


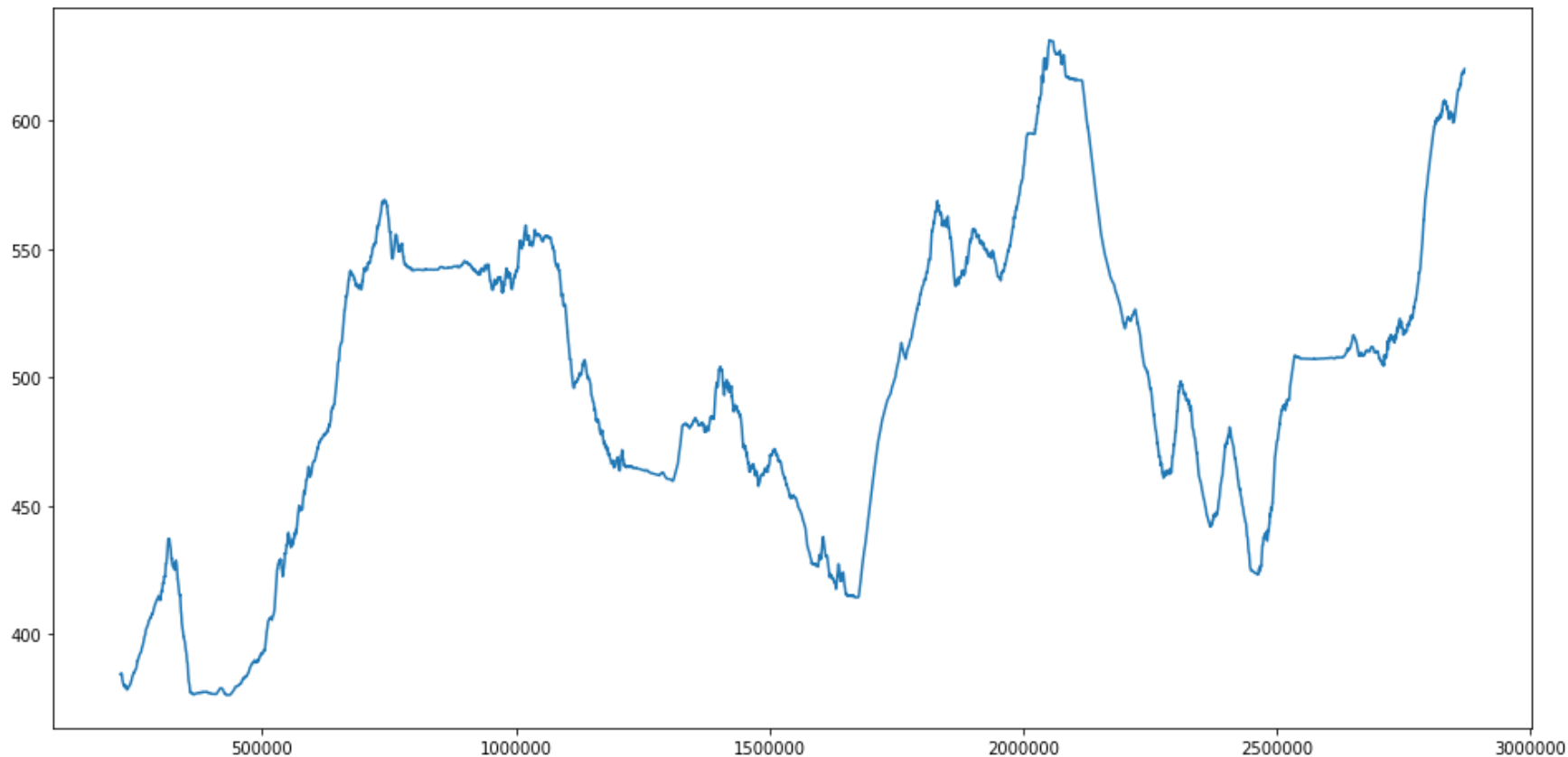
Time Series Decomposition

- `sys.dm_os_performance_counters` als Quelle

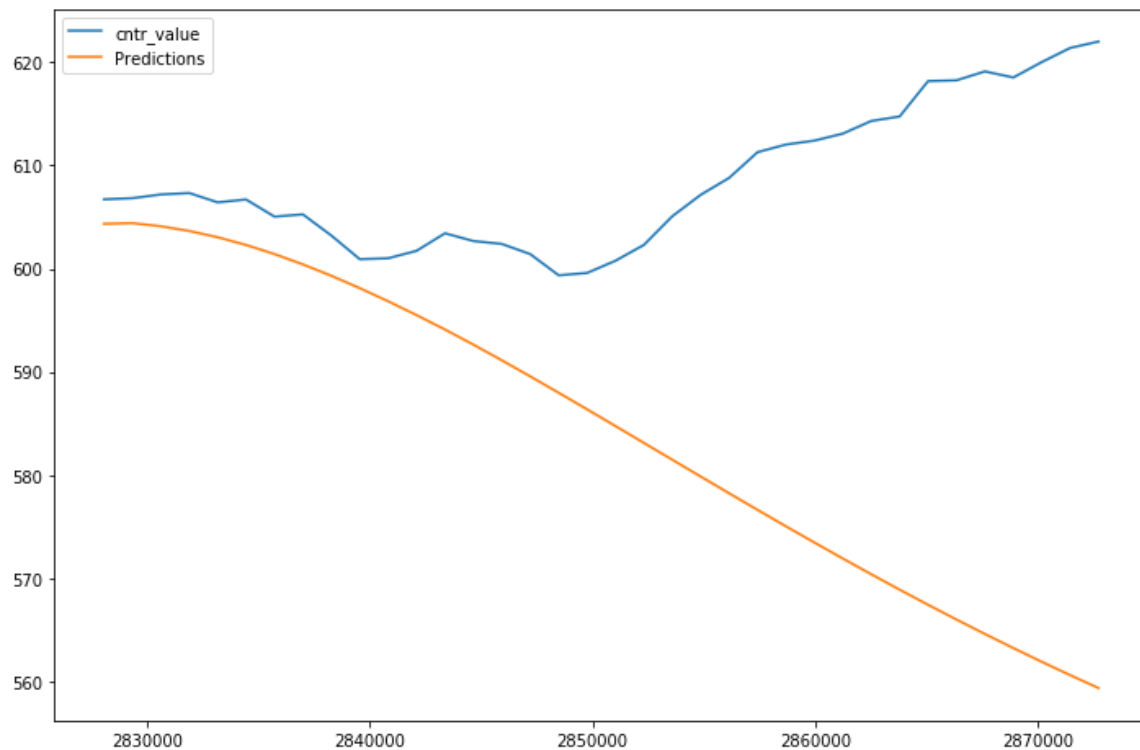








Time Series RNN Prediction



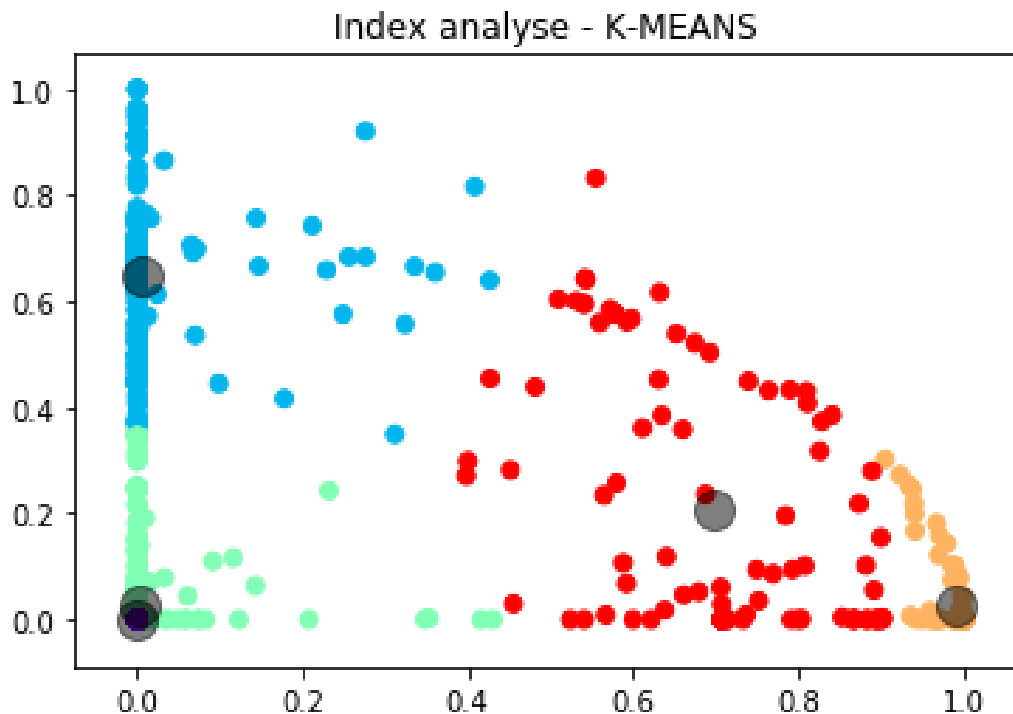


Unsupervised Learning

- `sys.dm_db_index_operational_stats()` als Quelle

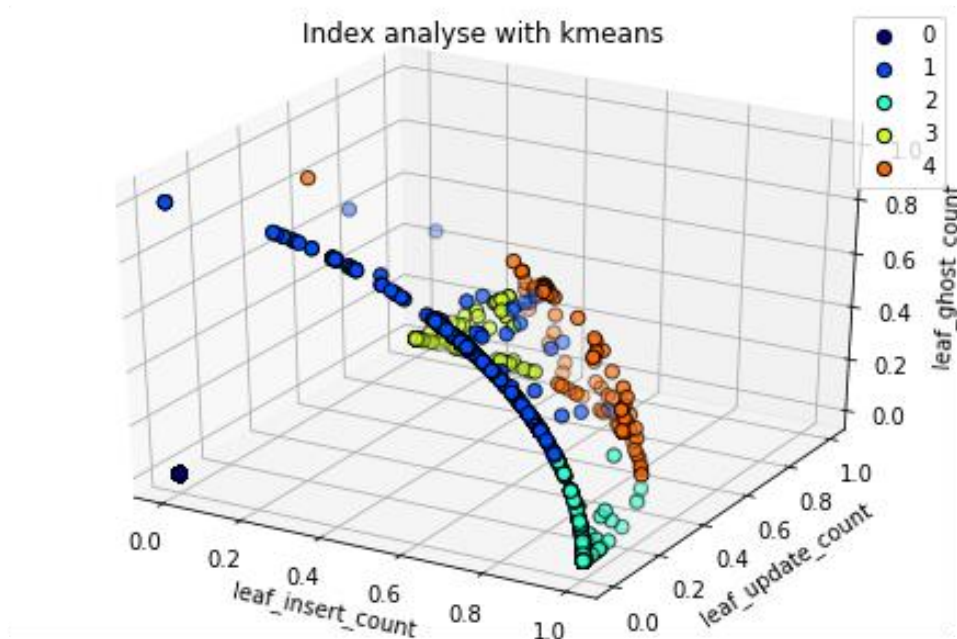


Index Analyse mit K-Means





Index Analyse mit K-Means





Learnings

- Machine Learning ist kein Projekt, sondern eine Reise
- Ihr braucht viel mehr als nur „das eine Modell“!
- Denkt früh an Eure ML-Pipelines! Wie bekomme ich das eigentlich später produktiv?
- Geht kreativ mit den Möglichkeiten um!



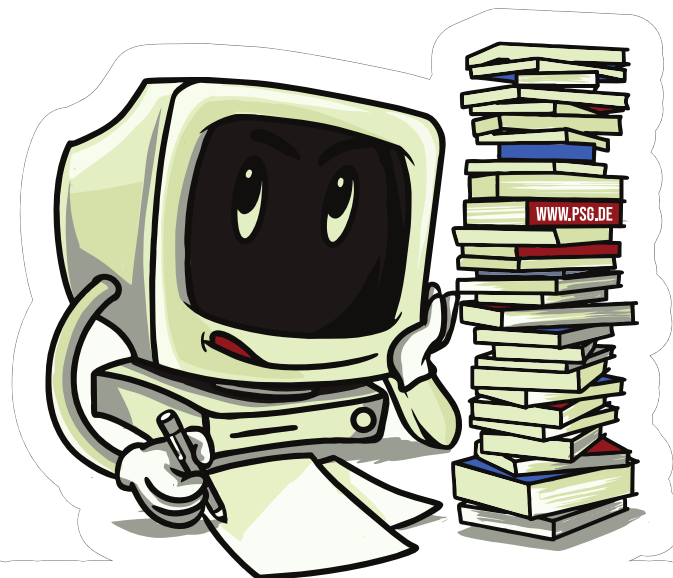
Thank you very much for your attention.
Vielen Dank für Eure Aufmerksamkeit.
und...



Feedback



Sticker hier vorne..



I BELIEVE IN MACHINE LEARNING!