



TeamViewer

Expelliarmus!

How to disarm competition with smart data tricks

2nd BI, Analytics & Data Management Summit
Vienna | 22-23.10.2019

@FilipVitek, Director Data Science

Expelliarmus?

Beg your pardon?



Data team is powerful GUN



But guns have one important issue



Unless some companies
use it for ...



Avada Kedavra vs. Expelliarmus



Avada Kedavra

- **Agile fix, buying time to stay** ahead of opponent
- Does **not consume** too many resources
- Opponent does **not feel urge to revenge**
- **Does not humiliate** others, some cooperation with them still possible in future
- **Opens doors for other options**, if this one fails
- Forces you **to innovate** (can't play the same trick twice)

- **Takes a lot of effort & focus** to properly execute
- Often **aims at** clearly **beating** the opponent
- **Bets all on single card**
- If fails to work, it might back-fire, as **opponent seeks his chance to strike back**
- If it works, it **becomes addictive** as quick fix
- Often has **no follow-up plan** (after all opponent is assumed to be dead)

Expelliarmus



In next minutes let me show you ...

18 data tricks

... that are more like **Expelliarmus**.

Ready?

Ah, 1 more thing...



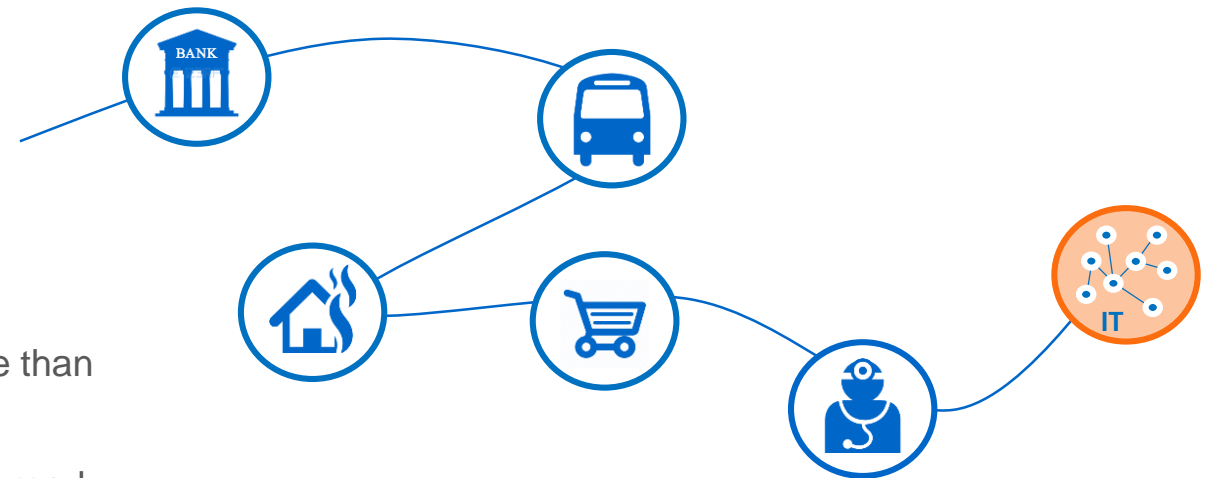
Who the hack is Filip Vitek ?



Mr. Filip Vitek

17 years building business strategies, **Data Science**, CRM systems development and BigData projects

Built **analytical units** in **6 different industries**, now working for **Teamviewer (IT)** :

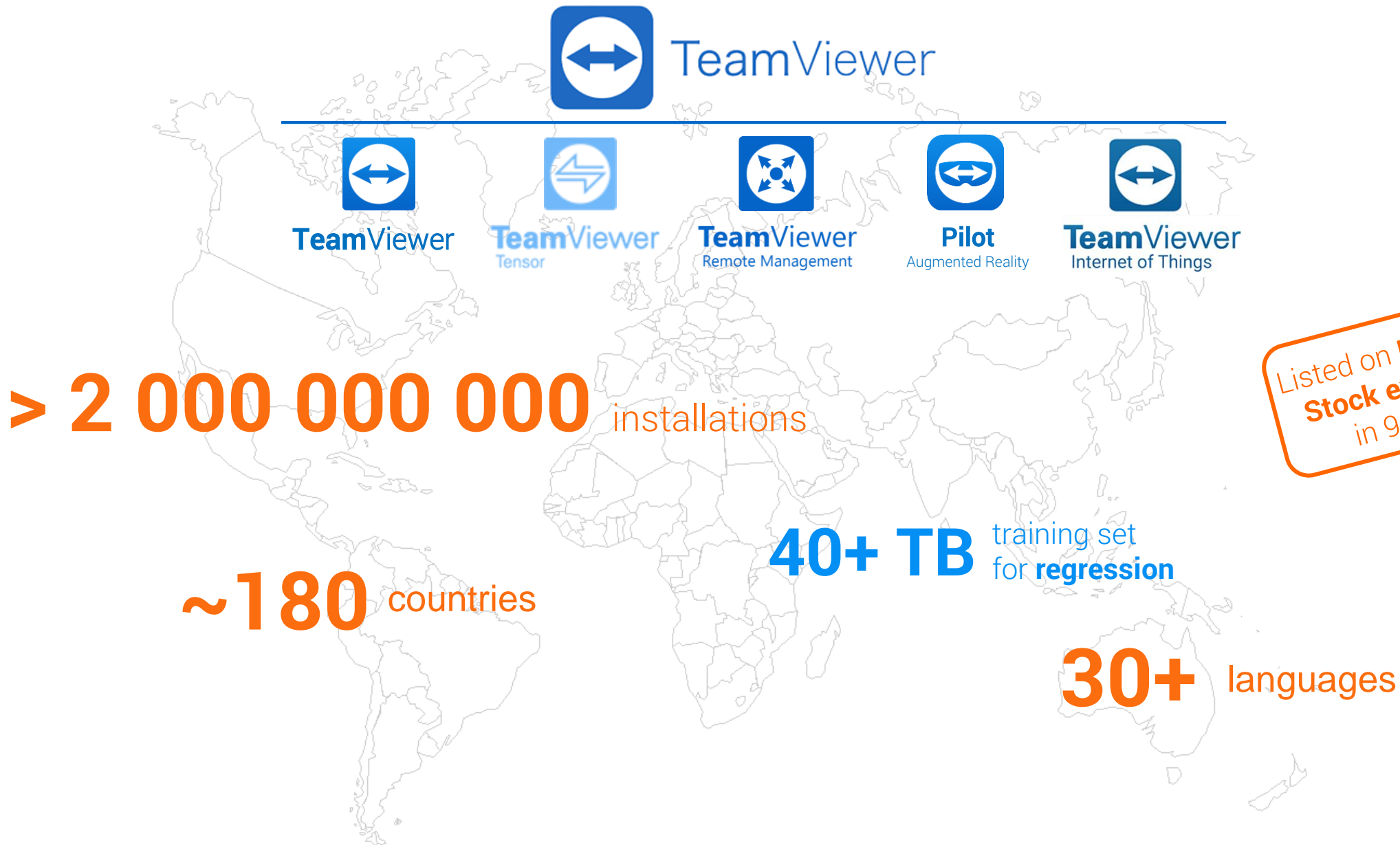


Data mining is my hobby and passion, I wrote more than

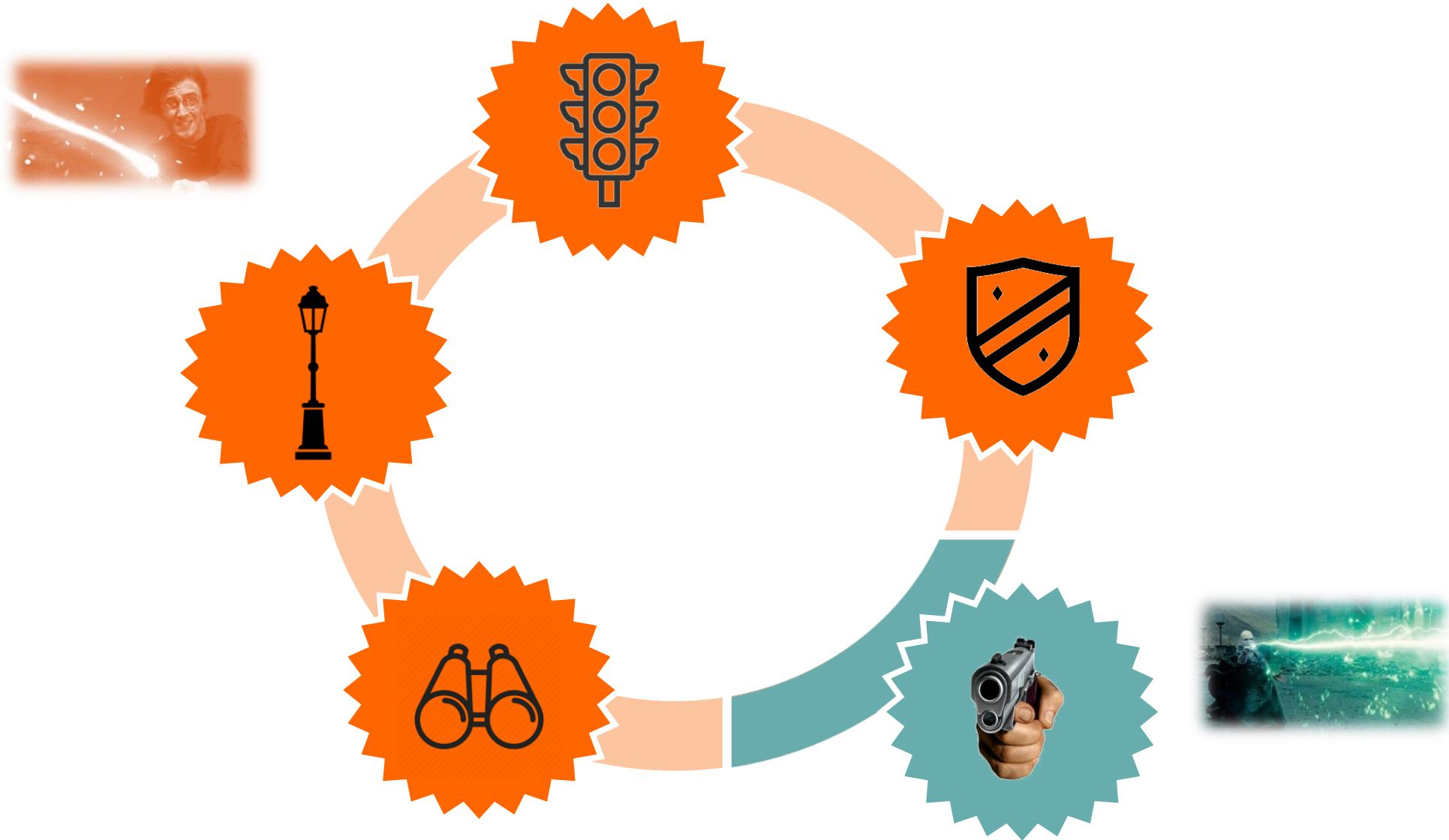
300+ expert blogs

If no time to go into details, I will leave you a link to read further on given topic.





There is so much more on Expelliarmus side



Collecting surprising external information

Data teams when faced with business problem have tendency to rely **MOSTLY** on already available data ... while there are always publicly available & free information. All you need to do is grab them.

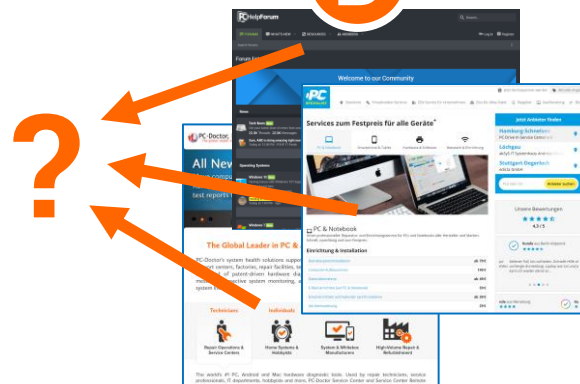
A Product & Price Intelligence



2x per day Web crawler of all products of competing e-shops

- Same day notification on any product added/delisted
- Complete history of any price changes of the products
- Comprehensive history of all promotion campaigns
- Analysis of the way that competitors describe the products

B Who are our customers?



x00 000s customers client base
What industry are they IN?

- 7 lines of code in Python
- Downloaded all web pages of B2B customers, isolated key words
- Detecting both industry as well as size of the company
- Application can now profile any new, prospective customer

C Who stole our clients?

Very competitive MTPL insurance
Who is stealing us most?
What is their strategy?



- Public register of the car insurance (for accidents)
- Ask on all cars that you lost in previous year
- Find out which company attacked which segment

Don't accept samples, strive to get FULL INFO



Point the light focus onto **your own pipes** ...

Google Analytics, SEO, Benchmarking, Social media analysis, Macro trends ...
are all following GUN principle ...

... and fail to document and tune very own
(often customer facing) processes

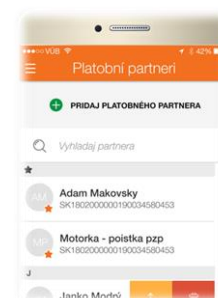
D *Darwinian analysis of own web*



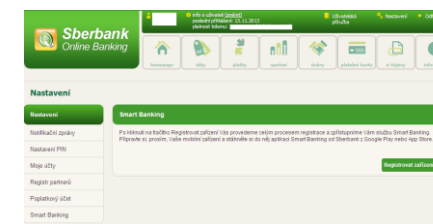
- Google analytics is useful for **analyzing funnel** on your page
- But drags your attention from **parts that do not work**
- If you have a **lot of content** (or even more content on the same url) you need to **apply Darwinian analysis**



E *Default value **per Planet**, not per User*



F *Collect **PROACTIVELY** events from your products*



Collecting session data & logs

- Understanding client's **intentions & frustrations**
- Which features are predominantly used by **which customer (groups)**
- Creates room for **very targeted messages**

We are living
Data Age, don't leave
**dark corners in your
company**



Discover your golden nuggets, your Data underdogs

Who will win the **car race** to nearest lights?



Has originally **other** informational role

Indicates **client behavior** [or its change]

☐ Data fields that are “**just identifiers**”

☐ **Contact & Transactional** data (IDs)

☐ No obvious relations as **champion challengers** (Joker cards)

☐ **Unusual** aspect of **usage**

☐ “Ryanair-like” data test

Data underdogs ... and their impact

Bank preference

(Online bank vs.
Postal bank)



Relationship proxy

(13333333 /xxxx
13335333 /xxxx)

Social impact on other clients in portoflio

Jane.Angry@company.com
Martin.Neutral@company.com



Tone of voice

John.Warton.Dr@hotmail.com

Johny_geek@hotmail.com

Mobile operators (and their share)



VIP indicator
(0903 444 555)

Real examples of Unconventional Parameters used



How old are you, Bernard?

- Nothing like “National ID” for German insurance companies = they have no clue about age of customer
- Important for setting proper communication (web vs. call vs. paper letter)
- **First name + Region** predicting 92% accurately the decade when the customer was born

[cut/off point for approx. 25% Individuals]

Fee increase tolerance

- Fee increase sensitivity for retail bank
- In search for metric that would tell: How “lazy” user is?
- Limited space, banking feels very un-emotional
- **Lowest amount** ever withdrawn from the ATM

[worked surprisingly well, due to big coverage]



We are sorry, you broke up with your girlfriend

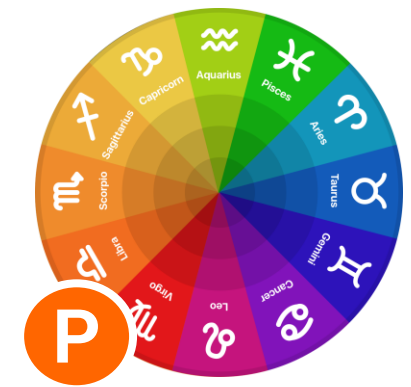
- Multidimensional data can be super powerful [own GDPR clause]
- Telco operators can aggregate about 4000+ parameters per each client
- **Only 10 parameters** needs Telco operator to detect that you broke up relationship with your partner

[pick-up time, rate of answer, night proximity, after air-flight outgoing message/call, ...]

Zodiac, are you kidding me?

- Probability to have car accident
- As Joker card for model
- Strong objection from Data Scientists: “*This is not serious work, we protest.*”
- Ended up as the **Second strongest** parameter in model.
- Later confirmed in **4 other countries** in same issue

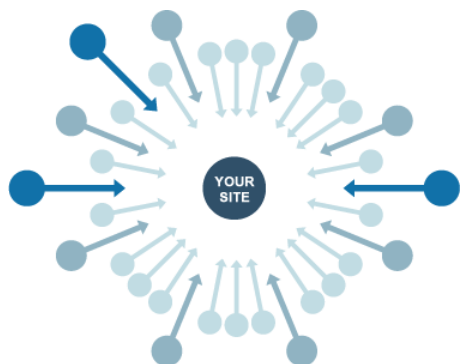
[I have a hypothesis why it works]



Others will fire their guns. Why not to get ready?

It always strikes me to realize, how eager the companies are to use data for offence, but how little are companies ready to shield their business with strong defense strategy coming from their data.

Q *Guarding access to your content*



- Detect which of the LINKs or IP addresses that are coming to your content are not desired ones
- Build black-list for traffic arriving from that sources, as they mean no good to you



If you know where to expect the attack, **strengthen your defense with data**

R *Price Decoy for the crawlers*



Being smart about traffic on your web page

- Built-in plug-in that is counting products visited from the same user/IP ...
- ... If is asking 1000+ product prices, engine starts to send random prices to deceive them
- Can work for content as well
- Don't forget to whitelist the price comparison sites or SEO engines you want to look good at



Quick summary of today's key points



We are living
Data Age, collect data on
all your internal processes,
don't leave **dark corners**
in your company

Don't accept samples,
strive to get FULL INFO, it
is often publicly available

Look into your databases for
your own data underdogs.
They are waiting for you.

If you know where to
expect others' attack,
don't wait vulnerable,
strengthen your
defense with data



Thanks for Your attention and I am ready to answer
YOUR QUESTIONS



www.TheMightyData.com

<http://mocnedata.sk/en/lets-keep-in-touch/>

Join the discussion



If you
have any
question
later ...

Mgr. Filip Vitek
Data Science Director
TeamViewer, Berlin



+ 49 1525 309 8505

filip.vitek@teamviewer.com

@



<https://sk.linkedin.com/in/vitekfilip>

@FilipVitek

