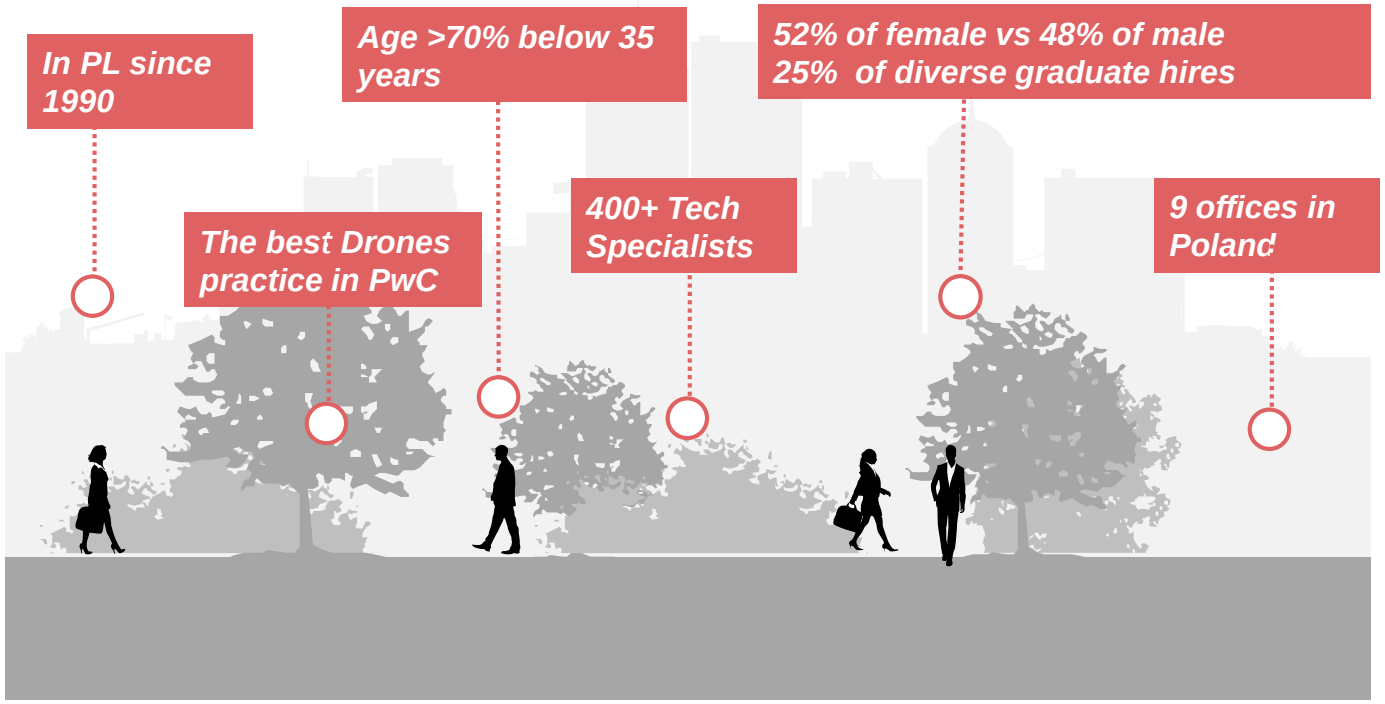
PwC

We build trust in society and solve important problems

pwc

PwC Poland key facts



Recruitment process

Graduates

www.kariera.pwc.pl

On-line tests

Assesment Centre / Interview

Internship offer

Experienced Hires

www.kariera.pwc.pl

1st Interview

2nd Interview

Job offer

www.pwc.com

Big Data Analytics Team Poland

pwc

Few words about me …

Education

- ***University of Warsaw***
  Mathematics
- ***Warsaw School of Economic***
  Quantitative Methods in Economics and Information Systems

Professional Experience

- ***Manager** in Big Data Analytics team*
- *6 years of experience in working with data for the biggest companies in Poland*
- *+20 data analytics projects for both FS and non FS sector*

Contact

- *pawel.marcinkowski@pl.pwc.com*
- *pl.linkedin.com/in/paweł-marcinkowski-a0638648*

What does Big Data mean?

Did you know...

Due to the lack of **analytical skills** companies analyse only 12% of data

Over 80% of data in enterprise consists of **unstructured data**

Every year data volumes explode by 40%

Poor data can cost businesses 20%-35% of their operating revenue

Big Data investments will account for nearly $40 Billion in 2015 alone

*To cope with data challenges, we present you **Big Data Analytics CEE Center of Excellence**, which is...*

... a developer of analytical solutions to any operational problems

...and a hub for strategic insight discovery

(by means of statistical modeling and optimization algorithms with advanced visualizations and real-time processing capabilities)

... an architect of analytical data marts or data lakes on Hadoop or SQL Server

...a provider of complex, tailored and generic apps that process small and big data to deliver simple insights on a specific topic

... and a qualified business understanding partner for all Advisory and x-LoS data-driven assignments

Our team contains currently over 50 specialist both data scientists and programmers

Big Data Analytics in numbers

Grade	*3 years ago*	*Now*
Senior Manager	1	2
Manager	0	4
Senior Associate/ Senior Data Scientist	1	10
Associate/ Data Scientist	2	15
Intern	0	11
Contractors	0	9
Total	**4**	**51**

Currently we are looking for

- ✓ Senior Data Scientist
- ✓ 2x Data Scientist/Intern

More info :

www.pwc.pl/pl/kariera/hello-world-data-analytics.html

Real-life use cases

Analytics services and products in general

Customer
- Enhance your targeting by personalization
  - Basket analysis
  - Recommender systems
- Segmentation
- Shop layout analysis
- Consumer Behaviour simulation (agent based modeling)
- Omni channel lifecycle optimization – predictive lifecycle modeling
- Churn modeling and prevention, loyalty analytics

Marketing
- Product Promotion Effectiveness
- Marketing- ROI (campaign analytics)
- Web analytics
- Sentiment analysis (social media, text mining)

Apps
- Credit-risk management system
- IFRS 9

Financial services
- Credit Risk modelling
- Stress testing
- Balance sheet simulation
- Portfolio valuation

Pricing
- Model-driven pricing
- Real-time pricing
- Targeted pricing strategies
- Pricing optimization
- Discount analytics
- Benchmarks

BI
- BI and visualization driven value discovery
- Management cockpits with BI capabilities
- Real-time analytics
- What-if analysis
- Fast Data Marts

Supply Chain
- Logistic and warehousing simulation
- Demand forecasting
- Stock and replenishment optimization
- Assortment planning

Geospatial
- Branch network analysis
- Network planning and reformatting
- Geo promotion analytics
- Potential analysis

Organization
- Analytics organization
- Customer-centricity: maturity diagnostic
- Analytics maturity assessment
- Analytics strategy

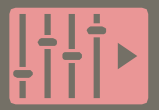
HR
- Productivity Diagnostic
- Workforce Demand & Sustainability
- Workforce Efficiency
- HR optimization
- Fraudulent behaviour detection

Marketing spending always under control

Understand the impact of effective campaigns...

In-depth analysis of past campaigns

Development of marketing plan



Real time product promotion effectiveness control

Define key drivers

Day of week

Channels

Marketing spending

Target group

Actions of competitors

... find the link between advertising and business performance

... which will boost ROI

Experience new dimension of your on-line presence

Web analytics

provides the means to measure

number of visits

cart abandonment rates

ad convection rates

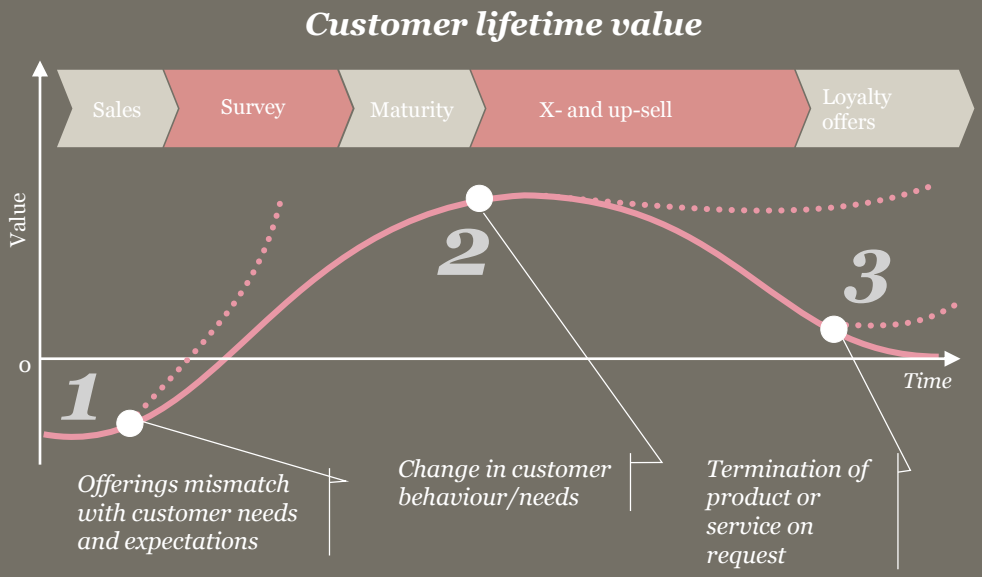
... and much more through click stream analysis maximize conversion rates

Tracking consumers behaviour in real-time across various social media platform using web crawlers and text mining to improve brand strength

Improve your customer segmentation with behavioural profiles based on on-line actions

Calculate customer lifetime value and optimize your spending

Customer expectations and priorities are constantly changing



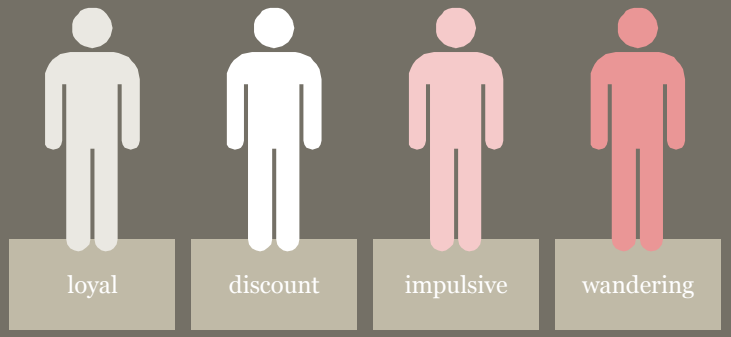
Marketing actions based on calculated expected revenue from customers over lifetime enhance effectively association with the company

Optimize marketing spending and provide the possibility of customers migration between channels according to customer lifetime value

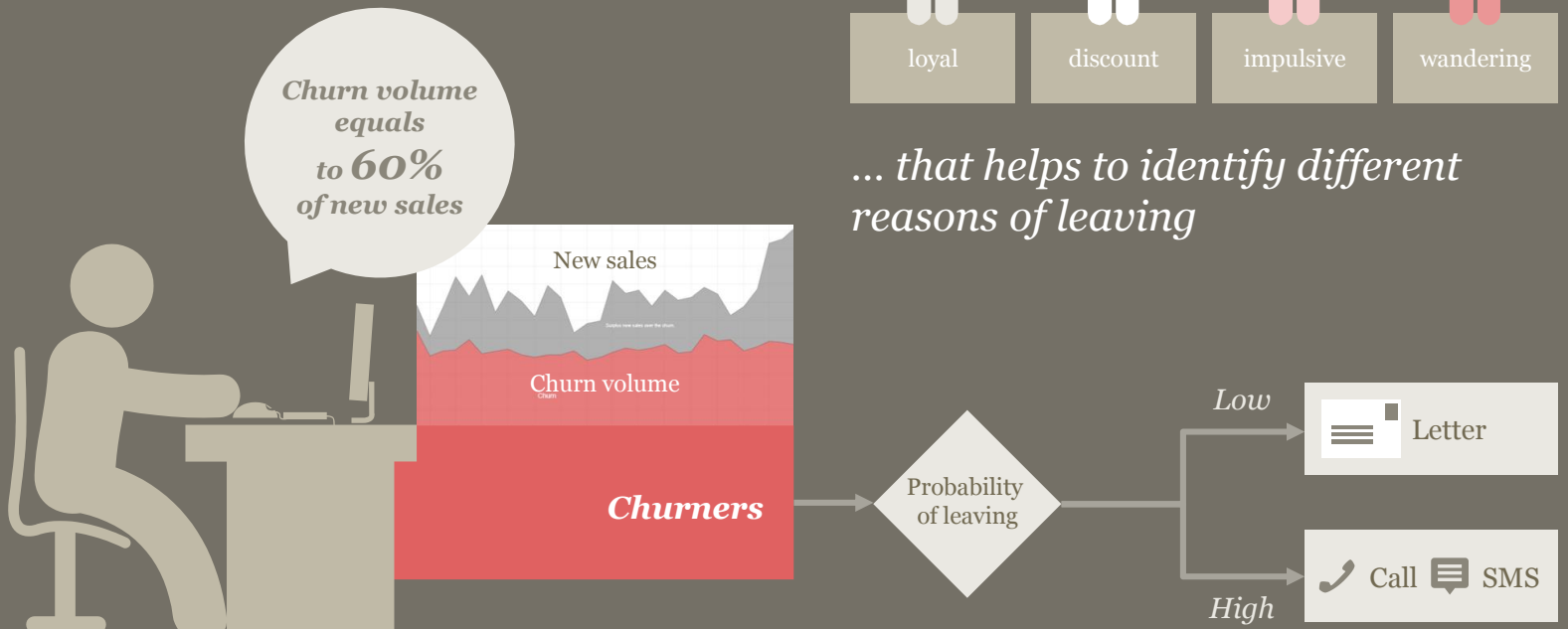
Detect churners and increase their loyalty

Big Data enable early detection of churners...

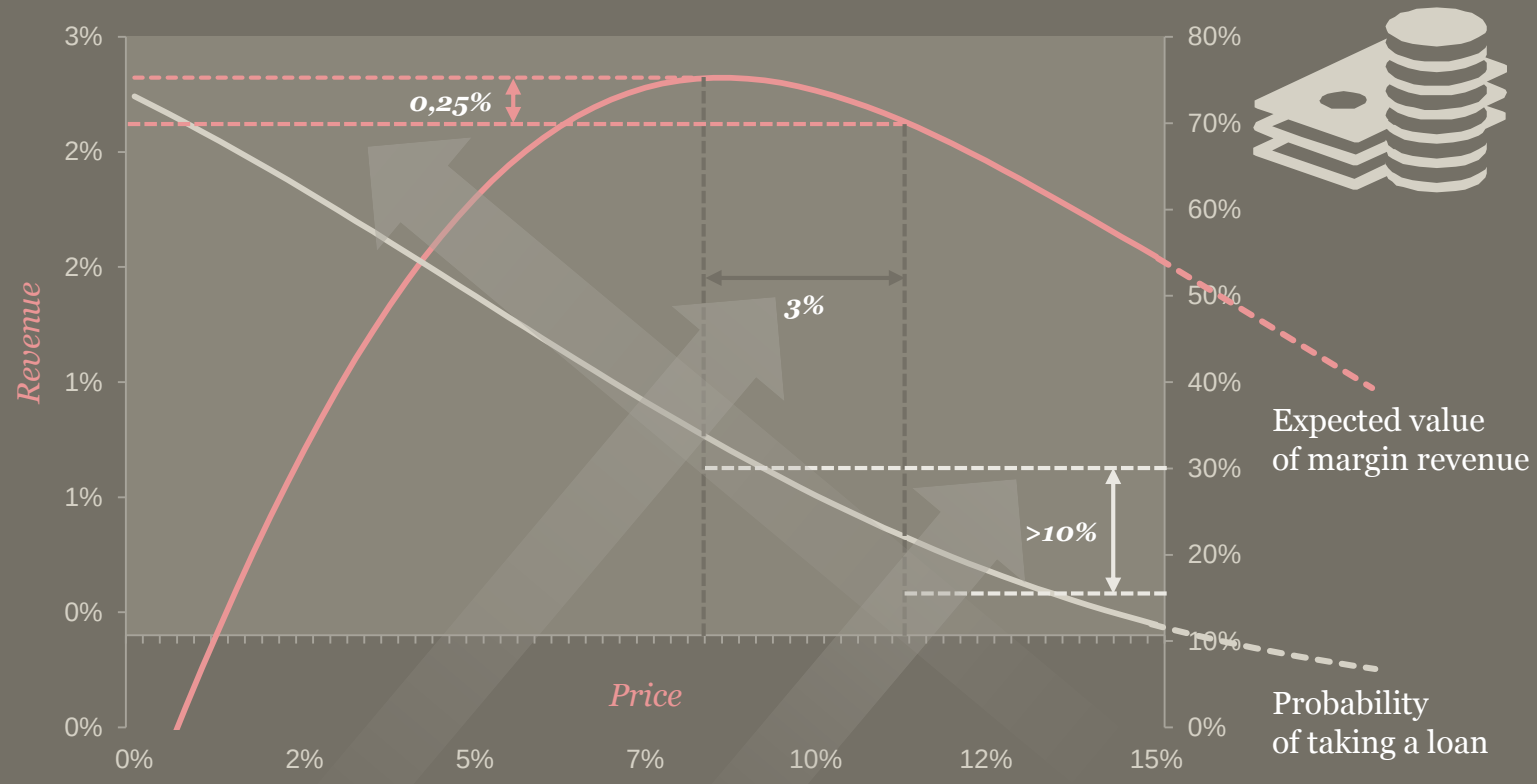
Enhance customer loyalty through unique plan of actions for each group



... that helps to identify different reasons of leaving



Find the ideal price for each customer to maximize revenue



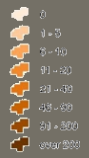
Introduction of flexible, model-driven pricing *can lead to attracting new customers.* *We are able to find the price that will maximize expected revenue.*

Real life application in a Polish bank delivered 20% margin uplift on a portfolio of newly granted loans

Find the most attractive locations

1 Capacity model

Summary of development directions for chain X



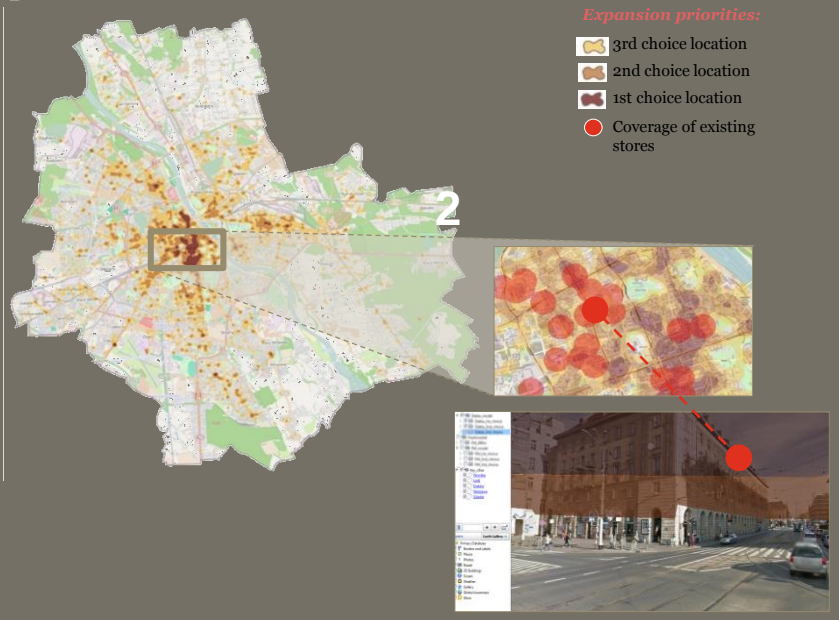
New openings potential [max. # of new stores]

2 Geospatial predictive analytics *increase accuracy of determination of best locations for network expansion*

It can be used to prioritize and aid search for new store locations even at street number level

Exemplary analysis from our project experience

Sample numbers

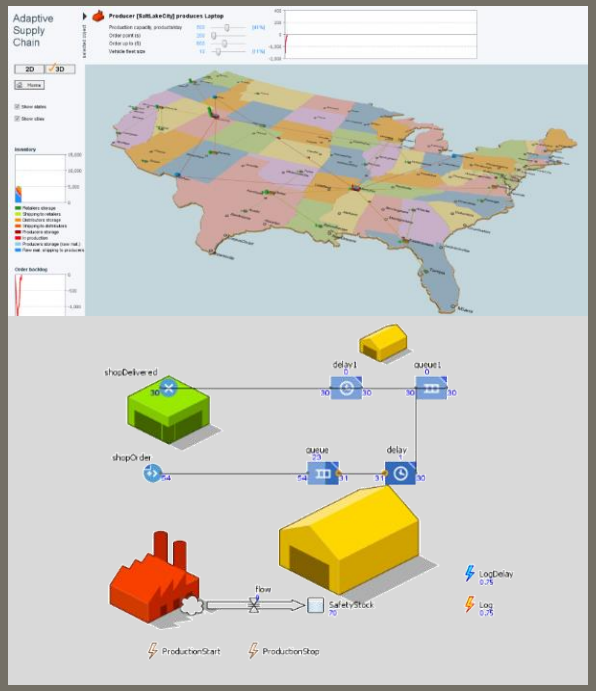


3 Business case

Up-to-date we have evaluated ca. 30 000 retail points and bank branches across CEE region

Optimize supply chain through logistic and warehousing simulation

Advanced tools enables simulation modeling of dynamic systems in manageable way



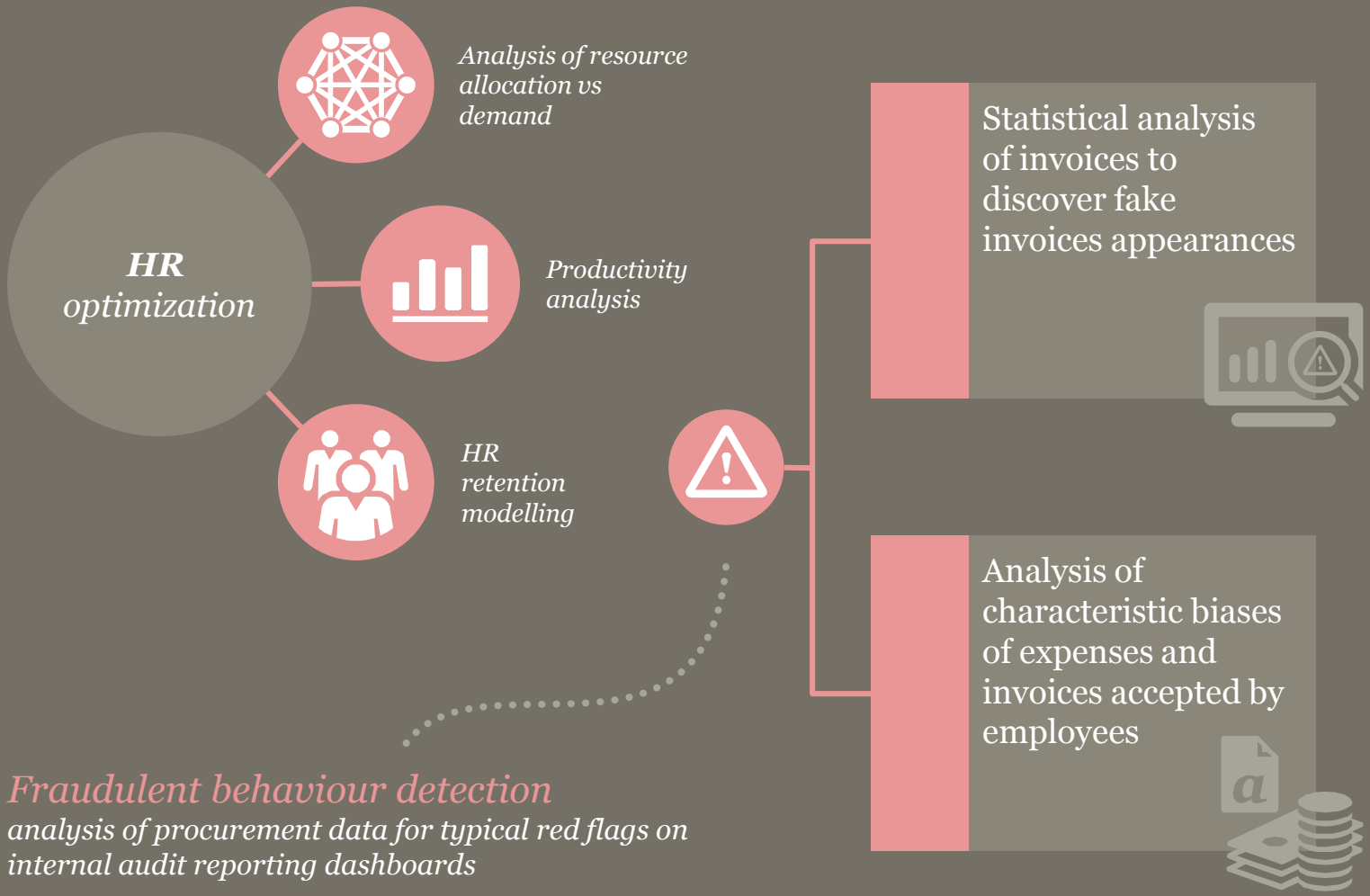
Enhance existing supply chain processes and ability to predict and react to events

Process evaluation...

field force optimization
minimize reaction time to failure

production, warehousing & logistic

Optimize HR through productivity assessment and fraudulent behaviours detection



Fraudulent behaviour detection

analysis of procurement data for typical red flags on internal audit reporting dashboards

Business Intelligence and visualization driven value discovery

Spot patterns and anomalies effectively

Make better decisions thanks to actionable insight

Data

Social media

Www, product, sensors, apps

POIs/maps

3rd party data

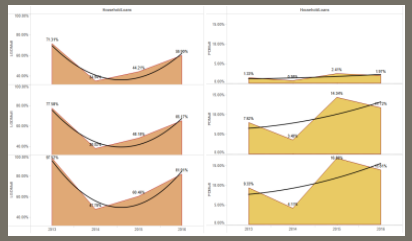
Risk modeling of credit and insurance enhanced by third party data sources

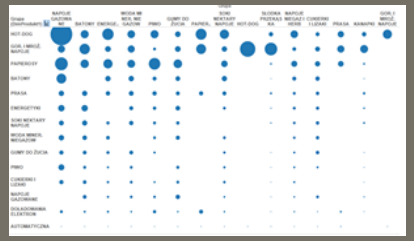
Comprehend risk concentration using satellite data, images and other third party data sources

Quantify exposure to risk

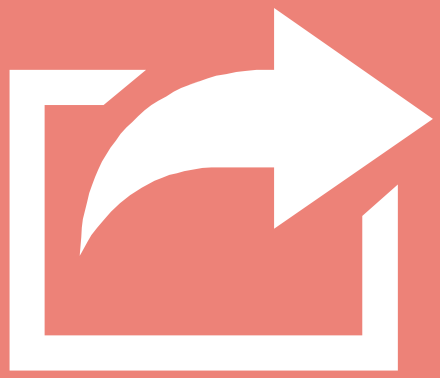
Identify leading risk indicators

Make better decisions based on data and minimize losses





Run whole calculations within minutes and reporting on almost live basis



What else do we have apart from slides?

Our toolbox

- Geographic data processing
- Visualization
- High-end geospatial solutions are created with QGIS+Python and PostGIS

- Best-in-class tool for various types of simulation, including agent-based

- Microsoft Visual Studio – integrated development environment (IDE) from Microsoft. It is used to **develop computer programs for Microsoft Windows, as well as websites, web applications and web services**
- Most IT solutions that are or will be created by our team are programmed in C#
- Our solutions are built on .NET platform

- Microsoft SQL Server – relational database management system developed by Microsoft
- Allows to **store, transform and share data with other users**
- Very high-performance database

- [R] – free software programming language and software environment for **statistical computing and graphics**
- It allows to perform **econometric and data mining computation**
- **We use SAS for compliance with typical banking and telco IT architecture**
- For fast prototyping we employ Python
- Genuinely Big Data solutions, processing large volumes of high-velocity data are built with Spark

- Tableau – **interactive data visualization** products focused on business intelligence
- Software that provides us a '**Wow**' **effect** during important client presentations
- Compatible with all most typical data sources
- **Top business intelligence tool** according to Gartner

- **State of the art BIG DATA architecture**
- **Parallel processing**
- Can handle all the data available– depending on the available storage
- Real-time analytics with Storm
- Numerous supporting technologies, including basic Pig&Hive, Spark, Splunk or Cassandra
- Cloudera distribution

Our techniques

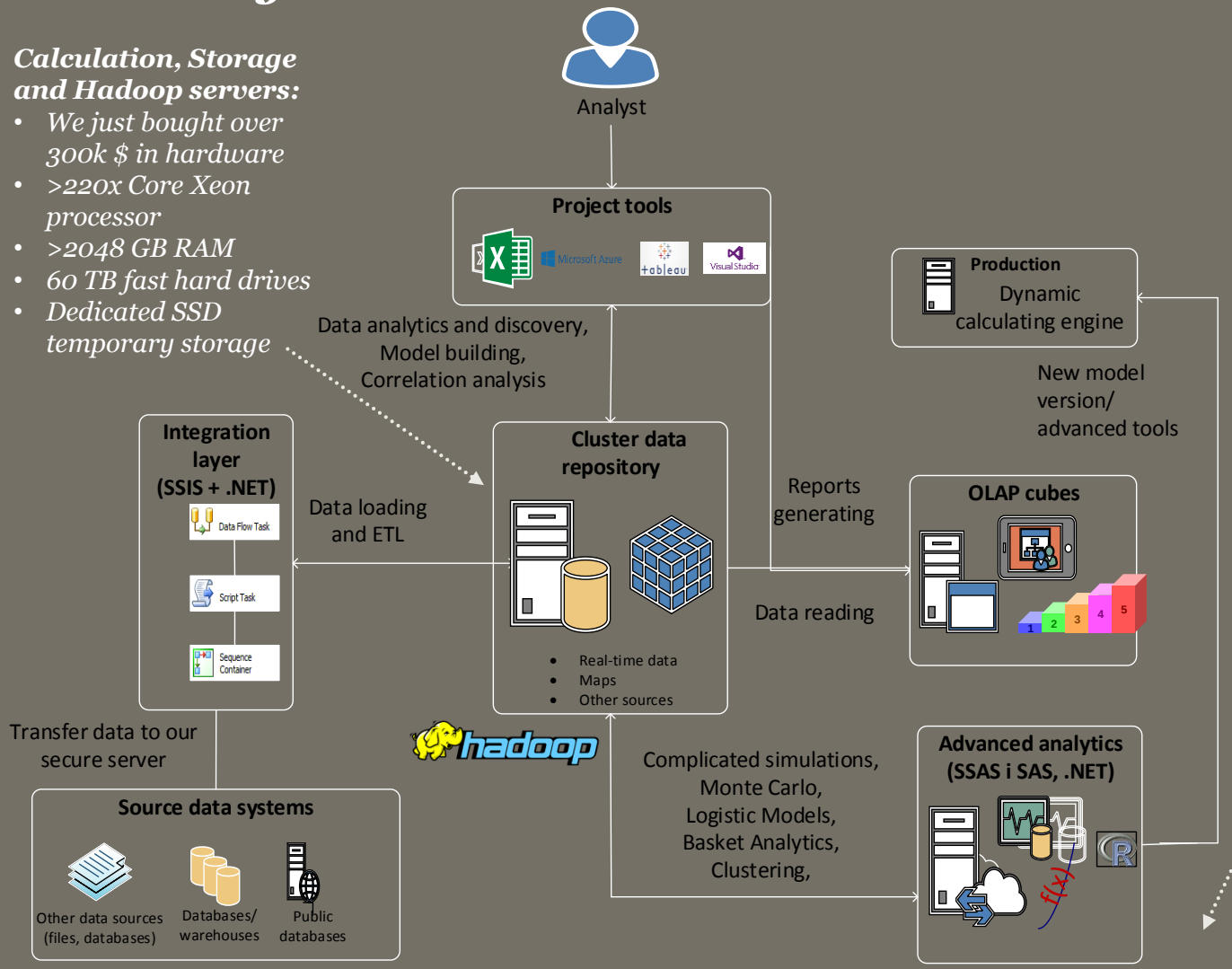
Our tools and techniques allow for implementation of models that are learning from data influx in real time and are able to self-calibrate, leaving client with end-to-end solution after project is completed

Text Mining	Parallel processing	
Text analytics	Programming: C#, Python, C++	
Real-Time processing	Advanced SQL	Spatial models for geoanalytics
Sentiment analysis		Logistic Regression
Supervised learning	A/B testing	
Unsupervised learning	Behavioral Clustering	Regularized regression (lasso, ridge)
	Graph analysis	
	Distribution estimation	Agent-based modeling
	Survey/Conjoint Analysis	System-dynamic modeling
Association rule learning	Optimization	Discrete event modeling
Monte-Carlo simulation	Segmentation/Clustering	
Markov Chain simulation		Time series analysis

We have a dedicated high-end infrastructure and process in place to explore data and prepare it for predictive modelling and analytics

Calculation, Storage and Hadoop servers:

- *We just bought over 300k $ in hardware*
- *>220x Core Xeon processor*
- *>2048 GB RAM*
- *60 TB fast hard drives*
- *Dedicated SSD temporary storage*



Heavy Quant processing abilities

- *A set of defined tools for advanced analytics*
- *Ability to profile, load, transform over 10 TB or more than 1billion rows*
- *Ability to program very sophisticated simulation methods and customer behavioural models using state-of-the-art. tools*

Thank you for your attention !