Reacfin

An actuarial engineering perspective: why actuarial science and IT skills must converge

12TH DECEMBER 2023

Xavier MARECHAL

CEO Reacfin and IA|BE qualified actuary

Expert in Non-Life and Health insurance (pricing, product development, reserving and risk management) and Data Science.

Reacfin Consulting

We offer consulting services in actuarial science & quantitative finance, including a.o. capital - portfolio - product - risk - and liquidity - management. We build our expertise on broad data science capacities.

Reaxii

By blending strong actuarial and financial business expertise with an in-depth understanding of cutting-edge IT technologies, Reaxii enables our clients to become more competitive and focus on their core business such as complex analysis, strategic decision-making and innovation.

Reacfin Academy

We share our knowledge with our clients. We offer a comprehensive learning platform, including on-site trainings, e-learning modules, webinars etc.

AGENDA

What are the new IT skills actuaries should develop?

Data Engineering
(ETL from various sources, understanding data pipelines)

Machine learning and AI
(AI, ML)

Automation and Scripting
(Python, PowerShell, automation tools)

DevOps Practices
(Collaboration between development and operations teams)

Blockchain technology
(Decentralized and secure transactions)

Cybersecurity awareness
(Risks and best practices)

Database management
(Database management systems, SQL)

Data security
(Encryption, access control and secure data storage)

Big Data Technologies
(Apache Hadoop, Spark)

IT Governance and Compliance
(Familiarity with frameworks)

Cloud computing
(AWS, Azure, Google Cloud)

API integration
(Data exchange)

SKILLS EVOLUTION FOR ACTUARIES

What are the IT-oriented skills actuaries should develop? ➔ My view ☺

Data Engineering (ETL from various sources, understanding data pipelines)	**Machine learning and AI** (AI, ML)	**Automation and Scripting** (Python, PowerShell, automation tools)
DevOps Practices (Collaboration between development and operations teams)	**Data Visualisation** (Tableau, PowerBI)	**Cybersecurity awareness** (Risks and best practices)
Database management (Database management systems, SQL)	**Advanced Excel skills** (Improve efficiency in data manipulation and analysis)	**Ethical considerations** (Data use, privacy and transparency)
IT Governance and Compliance (Familiarity with frameworks)	**Cloud computing** (AWS, Azure, Google Cloud)	**API integration** (Data exchange)

Reacfin

AGENDA

Skills evolution for actuaries

Creative data sourcing and management

Advanced predictive modelling

Upgraded coding, operationalization and reporting skills

Conclusion

Appendix

Reacfin

CREATIVE DATA SOURCING AND MANAGEMENT

Avoid the ""shit in, shit out"" effect

Data Engineering
(ETL from various sources, understanding data pipelines)

Data Visualisation
(Tableau, PowerBI)

Database management
(Database management systems, SQL)

Ethical considerations
(Data use, privacy and transparency)
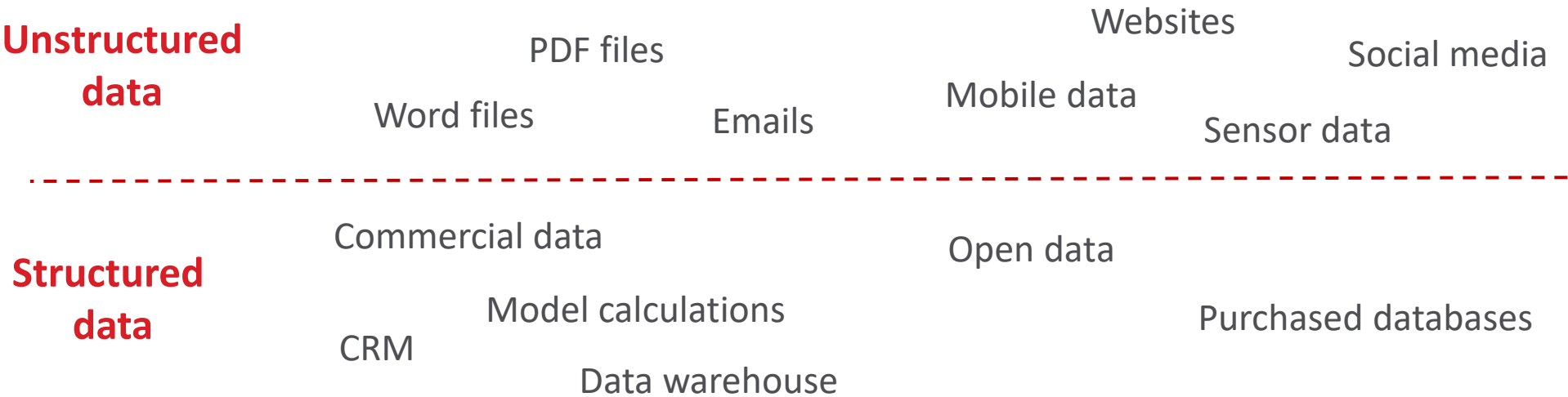
API integration
(Data exchange)

CREATIVE DATA SOURCING AND MANAGEMENT

There exist 2 main categories of data

- Structured data :

  organized and well characterized data that are easy to use because they are well identified.
  - E.g. insurer's policies and claims data

- Unstructured data:

  non-organized data not easy to manipulate and which require much preparation (everything else).

80%

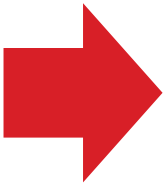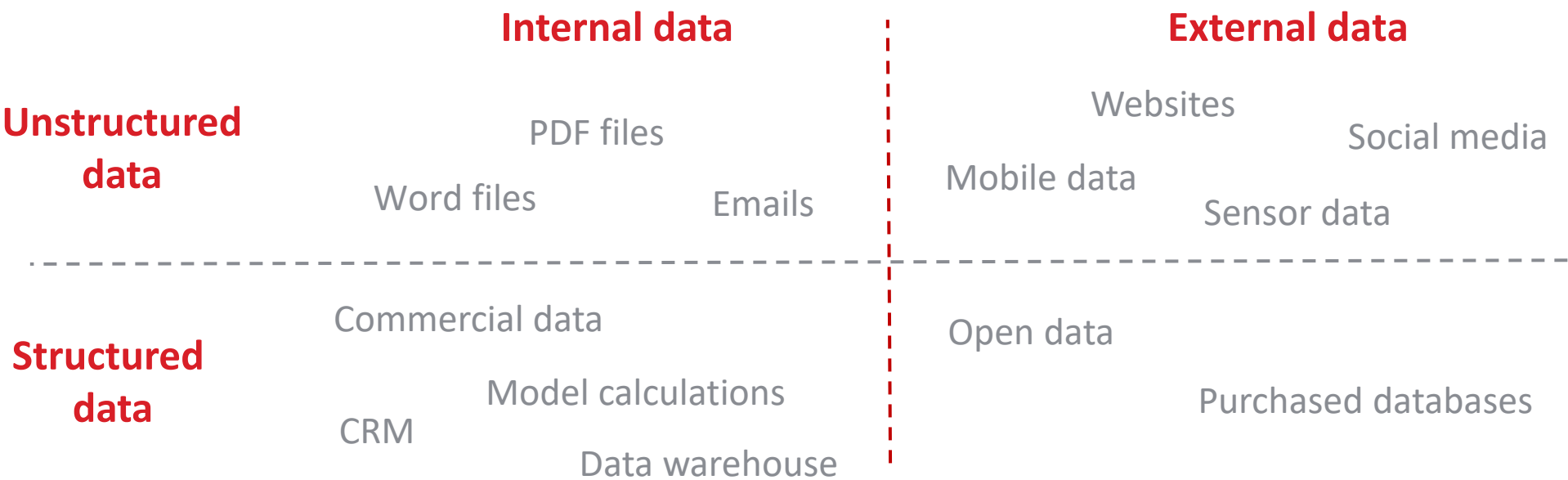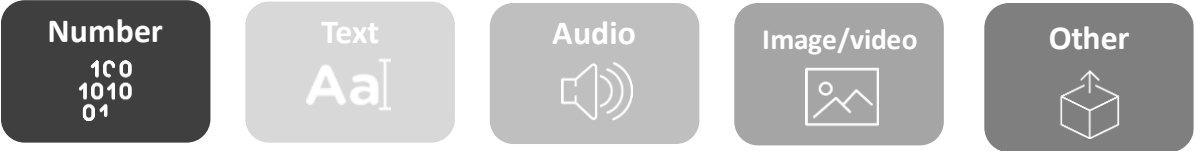Of business information are unstructured

Unstructured data

PDF files

Websites

Social media

Word files

Emails

Mobile data

Sensor data

- -

Structured data

Commercial data

Open data

Model calculations

Purchased databases

CRM

Data warehouse

CREATIVE DATA SOURCING AND MANAGEMENT

Different sources and types of information

- Numerous sources of internal or external data

- Data type is different from one content to another

| Number 1C0 1010 01 | Text Aa | Audio | Image/video | Other |

Increasing complexity to collect and manage unstructured data

	Internal data	**External data**
Unstructured data	PDF files Word files · Emails	Websites Mobile data · Social media Sensor data
Structured data	Commercial data Model calculations CRM Data warehouse	Open data Purchased databases

Data Engineering
(ETL from various sources,
understanding data pipelines)

Database management
(Database management
systems, SQL)

API integration
(Data exchange)

© Reacfin – 2023

Data sourcing

- Additional data can be obtained through **many different sources** :

    1. **Scraping/parsing techniques:**
        – Extract information automatically from websites

    2. **Open data files:**
        – Structured datasets available to everyone

    3. **IoT sensor and API technologies:**
        – Connected objects and application programming interface

    4. **External data providers**

    5. **Look twice into your own unstructured data:**
        – Reveal hidden information from core data

Thanks to their business knowledge, actuaries should be leading the identification and collection of additional relevant data

CREATIVE DATA SOURCING AND MANAGEMENT

Data Visualisation
(Tableau, PowerBI)

Data Engineering
(ETL from various sources,
understanding data pipelines)

Data engineering

- Once additional data has been collected, new methods and algorithm allow to get the most out of it. Among others:

1. Statistics, ML and feature engineering:

Create structured features using initial data sources or charts to understand data

2. Text mining, NLP and LLM

Process of examining large collection of written resources and methods to perform linguistic analysis and/or generation

3. Image processing

Techniques to perform operations on images to enhance its content or extract information

Actuaries should understand and make the most of these data engineering techniques to complete their traditional databases

CREATIVE DATA SOURCING AND MANAGEMENT

External and new data can be used to enrich the existing database with new attributes/variables

- **2 different points of view**
  - Data Scientist and Actuaries: enrich the existing database with a set of features which will be used when calibrating the models. More variables should lead to better predictive models (! Let's avoid overfitting !)

  - Business users (e.g. underwriter and marketing teams): simplify the models' results interpretation and/or the processes
    - e.g. reduce forms sizes in underwriting ➔ Quick quote (see appendix)

- **What happens if too much data?**
  - Need for feature selection: keeping only the most relevant variables
  - Potential for feature engineering: creating new variables to solve our problem

Reacfin

CREATIVE DATA SOURCING AND MANAGEMENT

Feature Engineering

- Features Engineering is absolutely known and agreed to be key to success in applied machine learning.

- Features Engineering is a Representation Problem
  - o Machine learning algorithms learn a solution to a problem from sample data.
  - o In this context, feature engineering asks: what is the best representation of the sample data to learn a solution to your problem?

Age	Frequency
-	-
-	-
-	-
-	-

Reacfin

CREATIVE DATA SOURCING AND MANAGEMENT
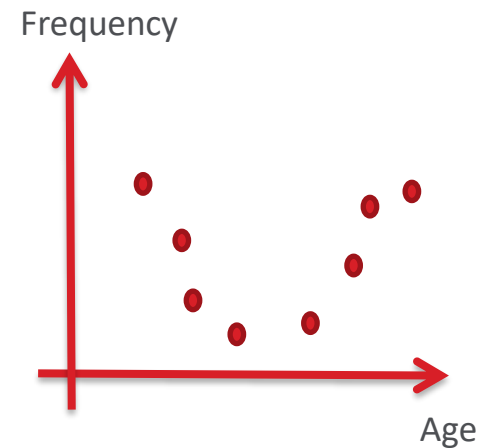
Feature Engineering vs Model Complexity

- The results you achieve are a factor of :
  - o the model you choose,
  - o the data you have available
  - o and the features you prepared.

- The better the features that you prepare and choose, the better the results you will achieve
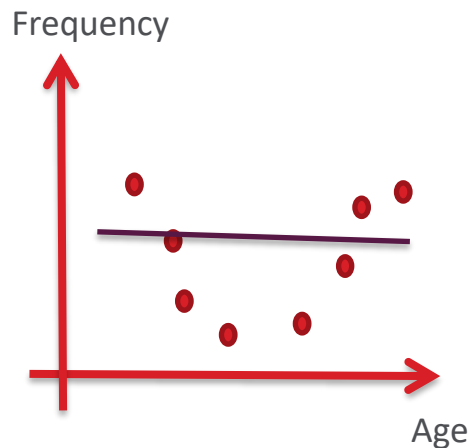
Linear (eg. GLM)	Non-linear (eg.GAM)	New Feature (eg Age^2)
Frequency / Age	Frequency / Age	Frequency / Age^2
Poor Model choice and without features engineering	More complex model and without features engineering	Simpler Model choice BUT features engineering
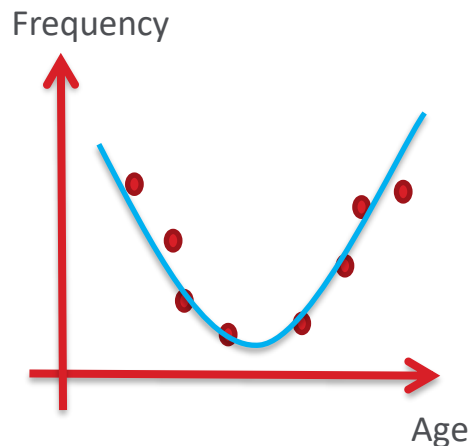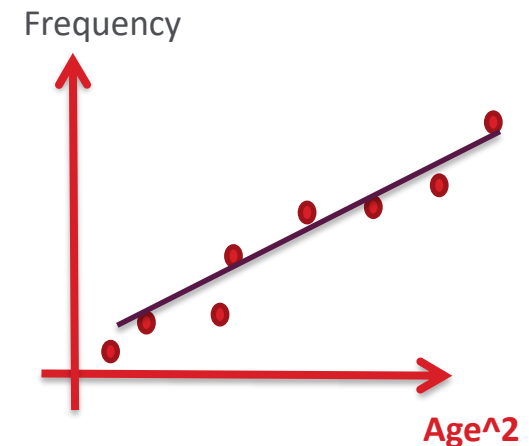
Reacfin

AGENDA

Skills evolution for actuaries

Creative data sourcing and management

Advanced predictive modelling

Upgraded coding and operationalization skills

Innovative reporting tools

Conclusion

Appendix

ADVANCED PREDICTIVE MODELLING

Machine learning offers new opportunities for predictive modelling but must be handled with care to avoid overfitting and/or black-box effect

Machine learning and AI
(AI, ML)

Data Visualisation
(Tableau, PowerBI)

	Machine learning	Statistical modeling
Limits the number of **assumptions**	+	-
Inference: Assessing the reliability of modeling assumptions	-	+
Prediction: ability to extrapolate future or unobserved realizations of a variable given other explanatory observations	+	-/+
"Big Data": ability to handle large sets of data both in terms of number of observations ("rows") or variables ("columns")	+	-
Human interactions: ability/need of incorporating material users ex-ante opinions (e.g. Expert Judgment)	-	+

- Results of Machine Learning algorithms will need careful attentions as they derive from automated procedures and could induce conclusions which do not match a business logic.
  - Interpretability is essential to avoid the black-box effect
- Another key challenge with Machine Learning is the risk of overfitting.

OVERFITTING

Overfitting deteriorates the predictive power of the model

The overfitting problem

- When modelling, we should be sensibilized with overfitting/lack of parcimony.

- It occurs when a statistical model describes random error or noise instead of the underlying relationship.

- The fact that the model fits our data well doesn't guarantee it will be a good fit to new data ➔ A good model is one that fits also well new data, i.e. that has a small predictive error
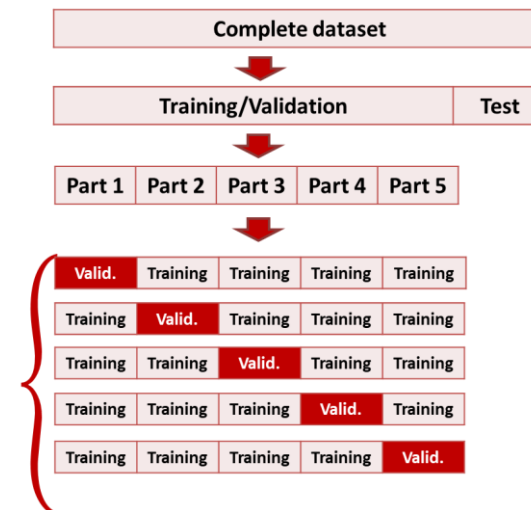

$$\theta_0 + \theta_1 x$$
High bias (underfit)

$$\theta_0 + \theta_1 x + \theta_2 x^2$$
"Just right"

$$\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$$
High variance (overfit)
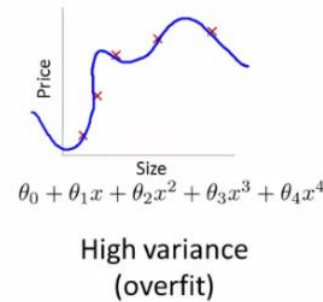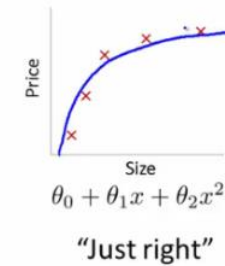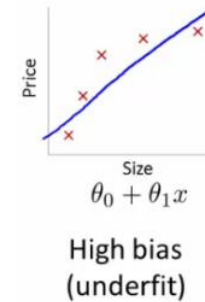
Potential solutions

Training set / Test set approach
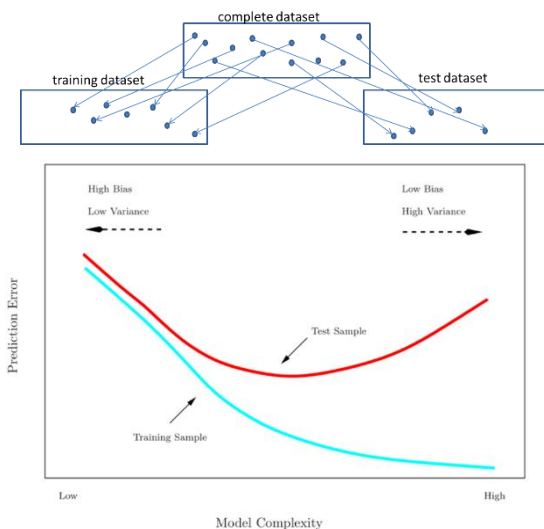
Cross-validation

Reacfin

SOME MACHINE LEARNING TECHNIQUES ARE BLACK BOXES AND INTERPRETATION OF THE RESULTS CAN BE QUITE DIFFICULT

Understanding the results of ML techniques is not easy. Therefore, using explainable ML techniques is a must!
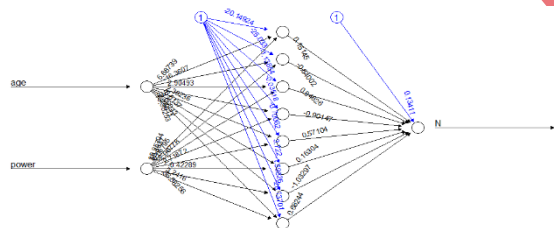
In the case of regression trees, understanding how the model predicts claims' cost or frequency values for new data points is not a problem, as it is very intuitive.
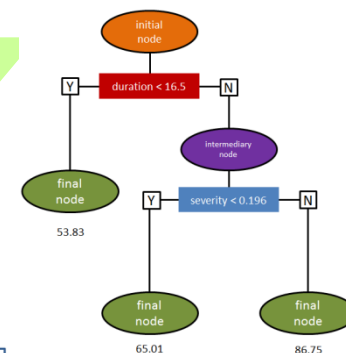
In the case of more complex methods such as Bagging and Random forests, even understanding how the model predicts values for new data points is rather difficult.
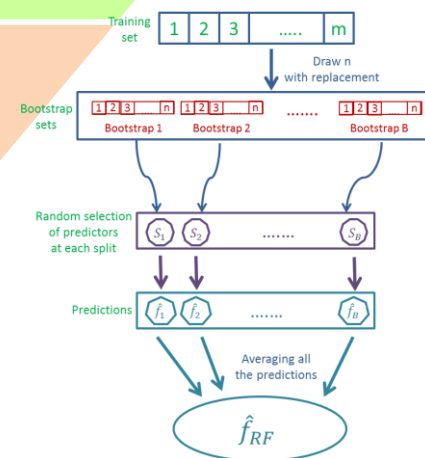
Things may be even worse for GBM and ANN.

Complexity

Interpretability

UNDERSTANDING THE RESULTS OF ML MODELS IS NEVERTHELESS KEY FOR SOUND BUSINESS DECISION-MAKING AS MANY STAKEHOLDERS USE THE RESULTS OF THE MODELS

Quant (Actuaries, data scientist,...)

- Able to understand the technical details
- Trust its outputs based on cross-validation, error measures and assesment plots

Other stakeholders

- Not necessarily « quantitative people »
- Should nevertheless understand and trust results to take decisions

Machine learning techniques usually improve predictive power but at the expense of a certain loss of interpretability ➔ Find trade-off between

| Predictive power | Capacity to understand the results | Ability to take sound decisions based on the results |

High-end questions
Who will use the results?   For what purpose?   With which impact?

AGENDA

Skills evolution for actuaries

Creative data sourcing and management

Advanced predictive modelling

Upgraded coding, operationalization and reporting skills

Conclusion

Appendix

UPGRADED CODING, OPERATIONALIZATION AND REPORTING SKILLS

Operationalisation of actuarial/financial computation engines is key for improving data-driven decisions

Automation and Scripting
(Python, PowerShell, automation tools)

DevOps Practices
(Collaboration between development and operations teams)

Data Visualisation
(Tableau, PowerBI)

Cybersecurity awareness
(Risks and best practices)

Advanced Excel skills
(Improve efficiency in data manipulation and analysis)

IT Governance and Compliance
(Familiarity with frameworks)

Cloud computing
(AWS, Azure, Google Cloud)

UPGRADED CODING, OPERATIONALIZATION AND REPORTING SKILLS

Pain points with (some) actuarial/financial processes

- Business experts (e.g., actuaries) sometimes **struggle with the efficiency** of actuarial/financial processes considering they must deal with

Data New data type – Quality – Privacy	**Functionality** Customization – Regression – Test	**Continuity** Compatibility – Model updates	**Interface** Interpretability – Friendliness – Portability
Performance Program efficiency – Time to production	**Security** Vulnerability – Network knowledge	**Compliance** Transparency – Doc IT standards – Regulation	**Operation** Environment – Architecture – Systems

- It may bring **painful consequences** for companies such as

Competitive disadvantage	Reputational risk	Over-costs
Critical operational issues	Resource shortage, dependency, attractivity	

UPGRADED CODING SKILLS

Improving coding governance

- Pain points identified
    - **No/bad versioning** of the codes developed
    - Difficult collaboration between developers and organization of tasks difficult
    - Difficulty to push the version in **production**
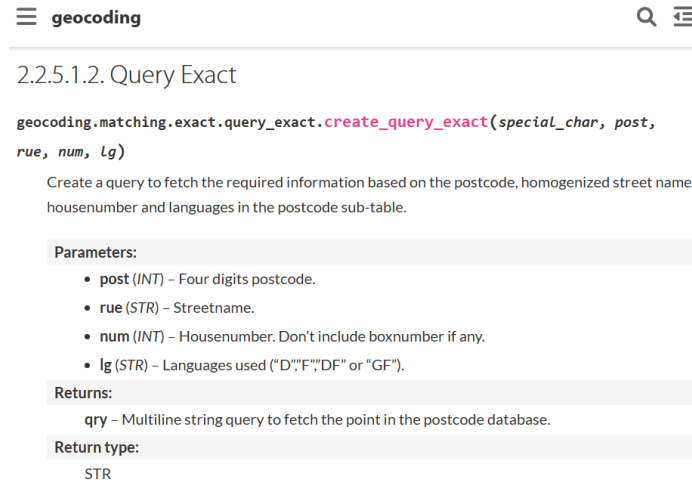
- Some solutions
    - **Git:** free and open-source distributed **version control** system and collaborative coding tool
    - **GitHub/GitLab**
        - Ideal to share and deploy code based on **Git**
        - **CI/CD platform** (Continuous Integration / Continuous Delivery).
    - **GitHub Copilot** is an AI coding assistant allowing suggestions in real-time
    - **GitHub Project** is an online tool which implements the **Kanban** and **Gantt** methodologies in a digital way
        - Other solutions for project management: Trello, Jira, Notion

UPGRADED CODING SKILLS
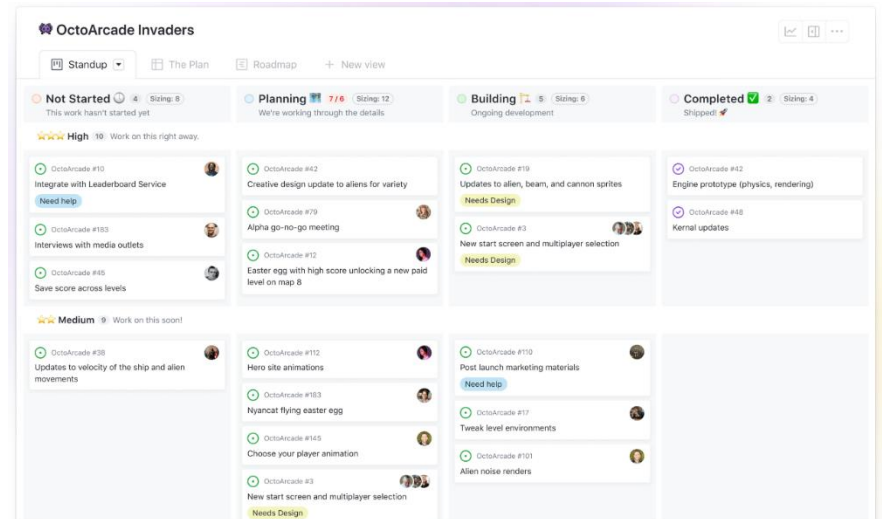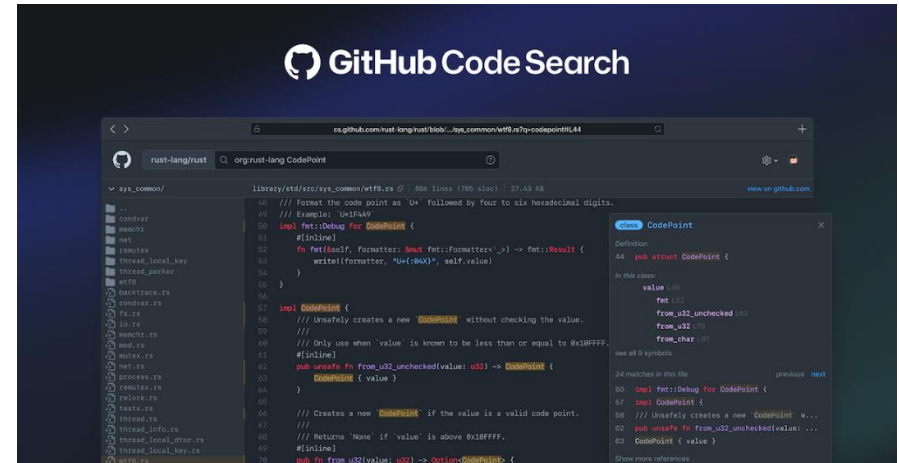
Decreasing key people risk with relevant documentation

- Pain points identified
  - Knowledge sharing and **business continuity** are key for actuarial teams
  - **Documentation** is not always up to standards
    - In particular the documentation linked to tools
    - It can create difficulties when changes/improvements are needed

- Some solutions
  - **Code to documentation** tools
  - **Sphinx** is a tool that makes it easy to create intelligent and beautiful documentation (in Python but also in C/C++)
  - Other tools: MkDocs, Doxygen

Developing robust prototypes and coding pipe-lines

- Pain points

  o Developing code in R Studio or another editor does not always allow to create a **clear project pipe-line** incorporating code, results and comments

- Some solutions

  o When prototyping, the cell-based approach of **Jupyter notebooks** is great.

    • They're great for showcasing your work. You can see both the code and the results.

    • It's easy to use other people's work as a starting point. You can run cell by cell to better get an understanding of what the code does.

  o Alternatives: R Markdown, CoCalc

➔ Using notebooks is probably a very good starting point if you want to standardize your coding pipe-line

© Reacfin – 2023
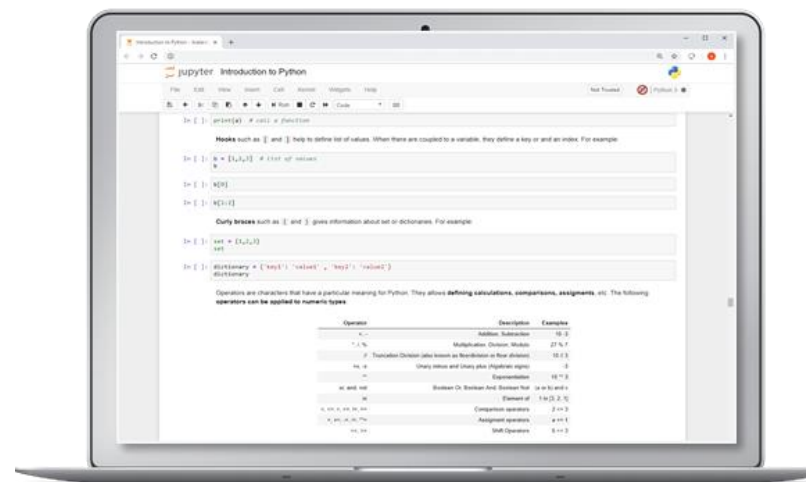
Improving durability and efficiency of applications

- Pain points
  - o Versions of open-source languages (e.g. R, Python) evolve over time, new packages are developed, or existing packages are adapted ➔ Difficult to **maintain an application over time**

- Some solutions
  - o **Docker** is a platform for developers to develop, deploy, and run applications with containers.
  - o A container is launched by running an image which is an executable package that includes everything needed to run an application--the code, a runtime, libraries, environment variables, and configuration files
  - o Other tools: Containerd, Kubernetes
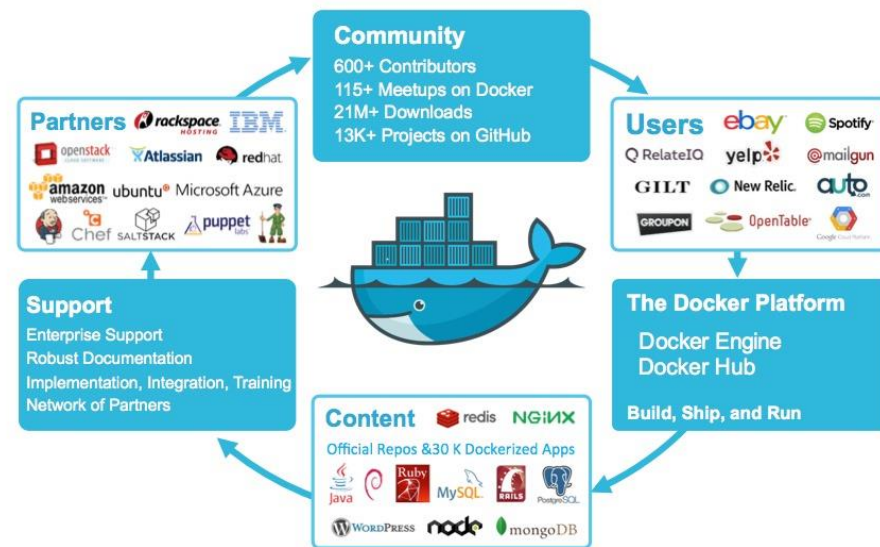
Reacfin

OPERATIONALIZATION AND REPORTING

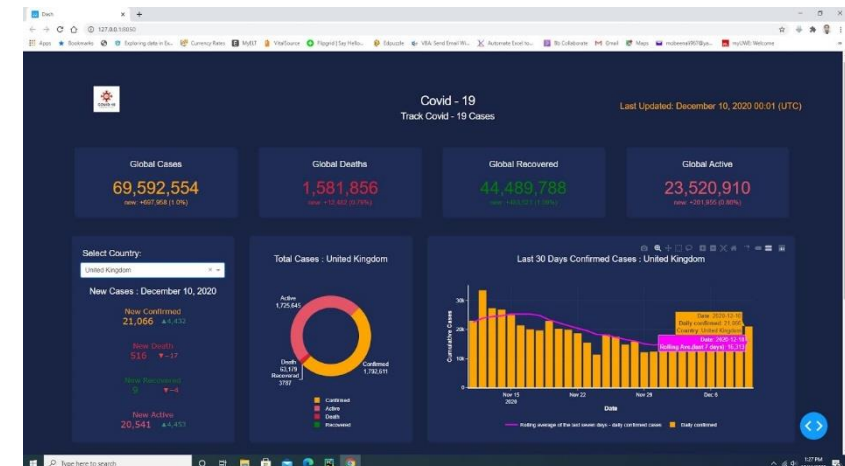**Visualisation is a must to communicate insight and improve story-telling**

- Pain points
  - **Communicating** results to other stakeholders is complicated as they might not be used to quantitative aspects
  - **Completeness and capacity to drill down** is not available in static reporting

- Some solutions
  - Developing dashboards, i.e. graphical user interfaces containing one or several data visualizations to monitor key performance indicators (KPIs)
  - **R Shiny** is a R package that allows to develop web application using R Studio (code editor)
  - The **shinydashboard** package is also really useful to create dashboards in Shiny.
  - **Dash** is a python framework created by plotly for creating interactive web applications
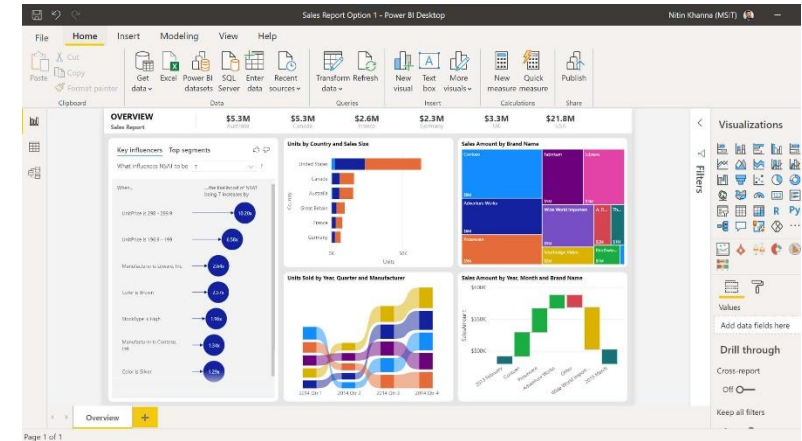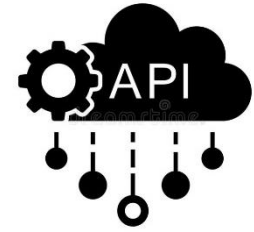
OPERATIONALIZATION AND REPORTING

Visualisation is a must to communicate insight and improve story-telling

- **Power BI** is a cloud-based business intelligence service suite by Microsoft.
  - It is used to convert raw data into meaningful information by using intuitive visualizations and tables. One can easily analyze data and make business decisions based on it.
  - It offers **drag-and-drop features** and **self-service capabilities**.

- A lot of **other BI tools** exist
  - Tableau, MicroStrategy, SAS BI, Qlik...

- Interesting visualization tools are also available online for **Excel**
  - In addition to the basic data visualization capabilities already embedded through charts, PivotTables or PivotCharts

OPERATIONALIZATION AND REPORTING

Application Programming Interface (API)

- APIs is an entry point to access assets (data, code, services,…) of a provider

- APIs have many advantages
  - Automation and efficiency
  - Integration and simplified access
  - Adaptation and possibility of customization
  - Easiness to share conent
  - Secure access to assets

- Examples of use in insurance
  - Digitalisation of operations
  - Increased efficiency in claims management (e.g. upload photos)
  - Developping a partners' ecosystem (aggregators, embedded insurance,…)

Reacfin

CONCLUSION

- Navigating the technological world is not easy as it is **evolving very fast**

- But there is a sweet spot for **actuarial engineers** between business and IT to develop robust actuarial and financial computation engines improving the decision-making process

- Thanks to their **business knowledge**, actuaries will be better placed than any profession to help insurance companies thrive should they **improve their technological skills**
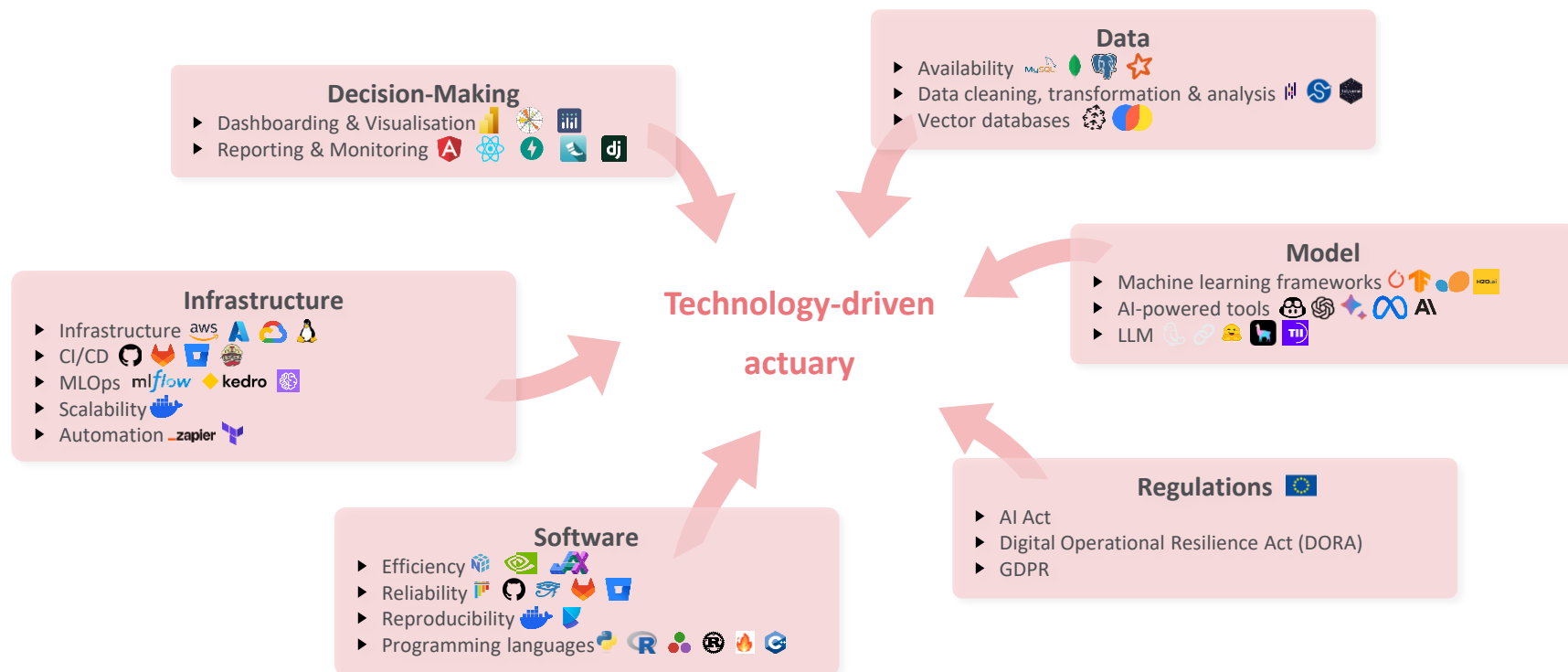
Thank you !

Do you have questions ?

About us

We develop, in partnership with our clients, actuarial & quantitative financial solutions, building on strong data analytics, while securing full transparency and integral knowledge transfer.

Reacfin

Place de l'Université 25

B-1348 Louvain-la-Neuve (Belgium)

T   +32 (0) 10 68 86 07

www.reacfin.com

Xavier Maréchal

CEO

M +32 497 48 98 48

xavier.marechal@reacfin.com

AGENDA

Skills evolution for actuaries

Creative data sourcing and management

Advanced predictive modelling

Upgraded coding, operationalization and reporting skills

Conclusion

Appendix

SKILLS EVOLUTION FOR ACTUARIES

What are the new skills actuaries should develop?

Economic and Financial Modeling (Assess impact of economic trends on risk)	**Data Sciences and analytics** (Visualisation, ML, predictive analytics)	**ESG knowledge** (Environmental, Social and Governance)
Programming skills (Python, R, SQL)	**Blockchain technology** (Decentralized and secure transactions)	**Cybersecurity awareness** (Risks and best practices)
Business acumen (Understanding broader business context, strategic mindset)	**Regulatory knowledge** (Impact of new laws and regulations on reporting and RM)	**Customer-Centric Focus** (Customer behavior and preferences)
Ethical considerations (Data use, privacy and transparency)	**Soft skills** (Collaboration, adaptability, problem-solving)	**Communication and presentation skills** (Compelling story with data)

Continuous learning

Reacfin

SKILLS EVOLUTION FOR ACTUARIES

What are the new hard skills actuaries should develop?

|		
Economic and Financial Modeling (Assess impact of economic trends on risk)	**Machine learning and predictive modeling** (More accurate predictions and improve risk assessment)	**Data Visualisation** (Tableau, PowerBI)
Programming and data analysis (Python, R, SQL)	**Blockchain technology** (Decentralized and secure transactions)	**Cybersecurity awareness** (Risks and best practices)
Advanced Excel skills (Improve efficiency in data manipulation and analysis)	**Regulatory knowledge** (Impact of new laws and regulations on reporting and RM)	
Data ethics and privacy (Data use, GDPR)		**Communication and presentation skills** (Compelling story with data)

Reacfin

CREATIVE DATA SOURCING AND MANAGEMENT

Case study: simplifying quoting process – Quick Quote by Generali

- In 2018, Generali Belgium launched **Quick Quote Car** a phone app that helps new customers to obtain a car insurance quote within 60 seconds

  o The app is available for insurance brokers who can obtain a quote with the following data

    • Picture of the car : Image recognition technologies help identify the license plate and the main characteristics of the car

    • Geolocation

    • Age of the driver
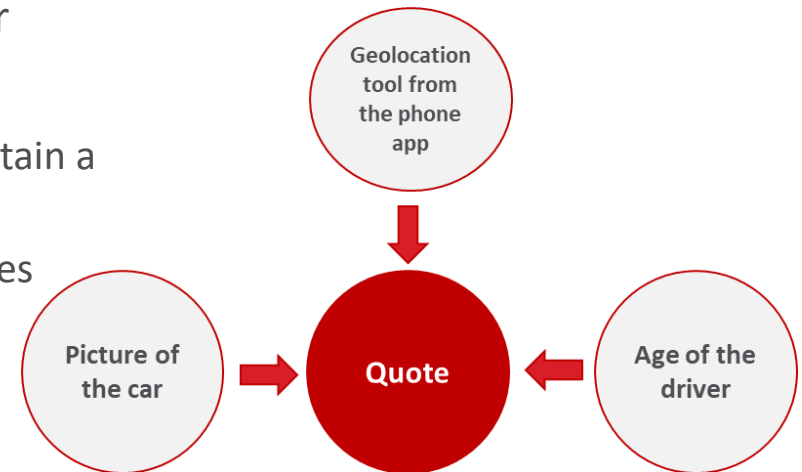
- Later in 2018, Generali also launched a similar app (Quick Quote Home) to help brokers calculate home insurance quote based on:

  o Pictures of the house :

    • Based on collaboration with a start-up specialized in real estate valuation (Rockestate), a 3D simulation of the house is made

  o Geolocation

NB: regulatory context has evolved since this initiative

CREATIVE DATA SOURCING AND MANAGEMENT

Data Engineering
(ETL from various sources,
understanding data pipelines)

© Reacfin – 2023

Features Engineering & Selection

- A feature is an attribute that is useful or meaningful to your problem

- Features Engineering is absolutely known and agreed to be key to success in applied machine learning

> *"At the end of the day, some machine learning projects succeed and some fail. What makes the difference? Easily the most important factor is the features used." — Prof. Pedro Domingos*
>
> *"Feature engineering is the process of transforming raw data into features that better represent the underlying problem to the predictive models, resulting in improved model accuracy on unseen data." — Dr. Jason Brownlee*

- Features Engineering is a Representation Problem
  - Machine learning algorithms learn a solution to a problem from sample data.
  - In this context, feature engineering asks: what is the best representation of the sample data to learn a solution to my problem?

Reacfin

CREATIVE DATA SOURCING AND MANAGEMENT

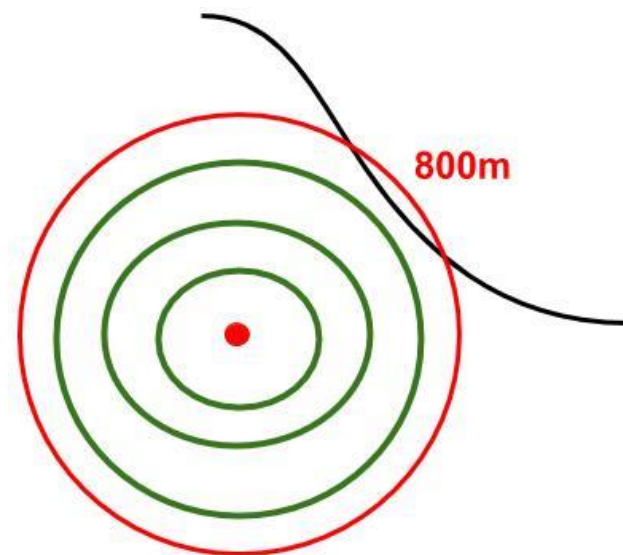**Case study: Features engineering in property theft insurance**

- Intuition : there is a correlation between the claims frequency and the distance from the highway
  - o Data available in the company : addresses
  - o Features Engineering : convert house addresses into distance from the highway
- Highway only?
  - o No, all the roads where the speed limitation is above 90km/h
- Determine the closest point to the highway in relation to the house.
  - o We need to know the location of the house on a map
  - o We need to know the location of the highway on a map
- Open Street Maps
  - o Gives the roads' longitude and latitude at different points.
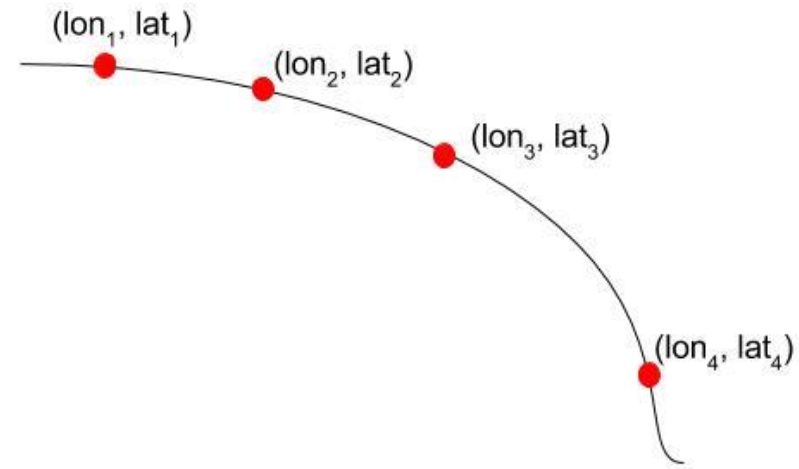
(lon_1, lat_1)

(lon_2, lat_2)

(lon_3, lat_3)

(lon_4, lat_4)

CREATIVE DATA SOURCING AND MANAGEMENT

Case study: Features engineering in property theft insurance

- If we want to find the distance from the house, we need the coordinates of the house.

- Google Maps Geocoding API
  - Geocoding is the process of converting addresses into geographic coordinates

- Google API is free but with slow performances:
  - 2,500 free requests per day
  - 50 requests per second (limitation of speed)
  - Enable pay-as-you-go billing to unlock higher quotas: $0.50 USD / 1000 additional requests, up to 100,000 daily.

- Find the distance between the house and the first road (above 90km/h).
  - We build a loop that checks if there is a road in a growing area (in a radius growing from 0 to 4000m with step of 200m)

800m

Reacfin

Reacfin

Place de l'Université 25
B-1348 Louvain-la-Neuve
www.reacfin.com