Capstone Project

AI Workflow - Capstone Project

Codename: AI_DS_IBM_Badge_Foundation

Paulo Calabria

Agenda

- Overview of the Business Problem
- Overview of the Project and Team (Stakeholders)
- Overview of the Data Science Solution
- Part 1 – Data Investigation
- Part 2 – Model Build and Selection (ML)
  - Logic Flow
  - Jupyter Notebooks
- Part 3 - Production

Business Understanding

Create a service that, at any point in time, will predict the revenue for the following month.

Service to have the ability to project revenue for a specific country.

Assumption: To keep the development time reasonable, the model will be limited to the ten countries with the most revenue

Project Objectives (Story)

- Currently, the managers are using their own methods to predict revenue, but they have come to feel that they are spending too much time on this and their lack of expertise in data science means their predictions are not as accurate as they would like.
- The management team expects to spend less time creating their own projection models, and they expect the new results to be more accurate.
- They have assured us that well-projected numbers will help stabilize staffing and budget projections which will have a beneficial ripple effect throughout the company.

Project Objectives

Testable Hypothesis

- H1
  - New Model will be faster and more accurate than the current model

Statakeholders

AAVAIL:

- *Management Team*

Data Science Team:

- *Paulo Cesar Pinto Calabria/Brazil/IBM – GTS – Data Scientist*

Overview of Data Science Solution

State the ideal data to address the business opportunity and clarify the rationale for needing specific data

Ideal data would be to have information per country about :

- purchases
- unique_invoices
- unique_streams
- total_views
- year_month
- Revenue

Part 1 Data Investigation

3. Create a python script to extract relevant data from multiple data sources, automating the process of data ingestion.

- From within a Python module there should be a function that reads in the data, attempts to catch common input errors and returns a feature matrix (NumPy array or Pandas DataFrame) that will subsequently be used as a starting point for EDA and modeling.

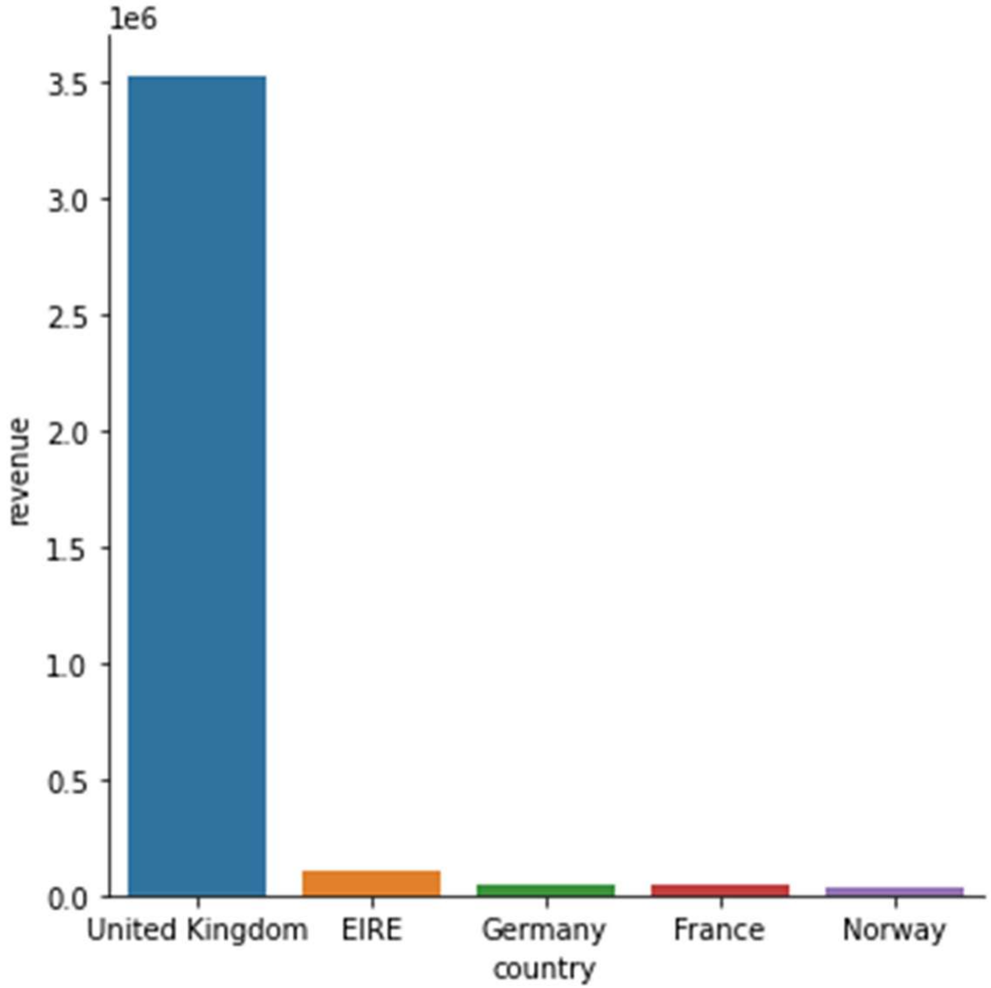
4. Investigate the relationship between the relevant data, the target and the business metric.

- Using the feature matrix and the tools abvailable to you through EDA spend some time to get to know the data.

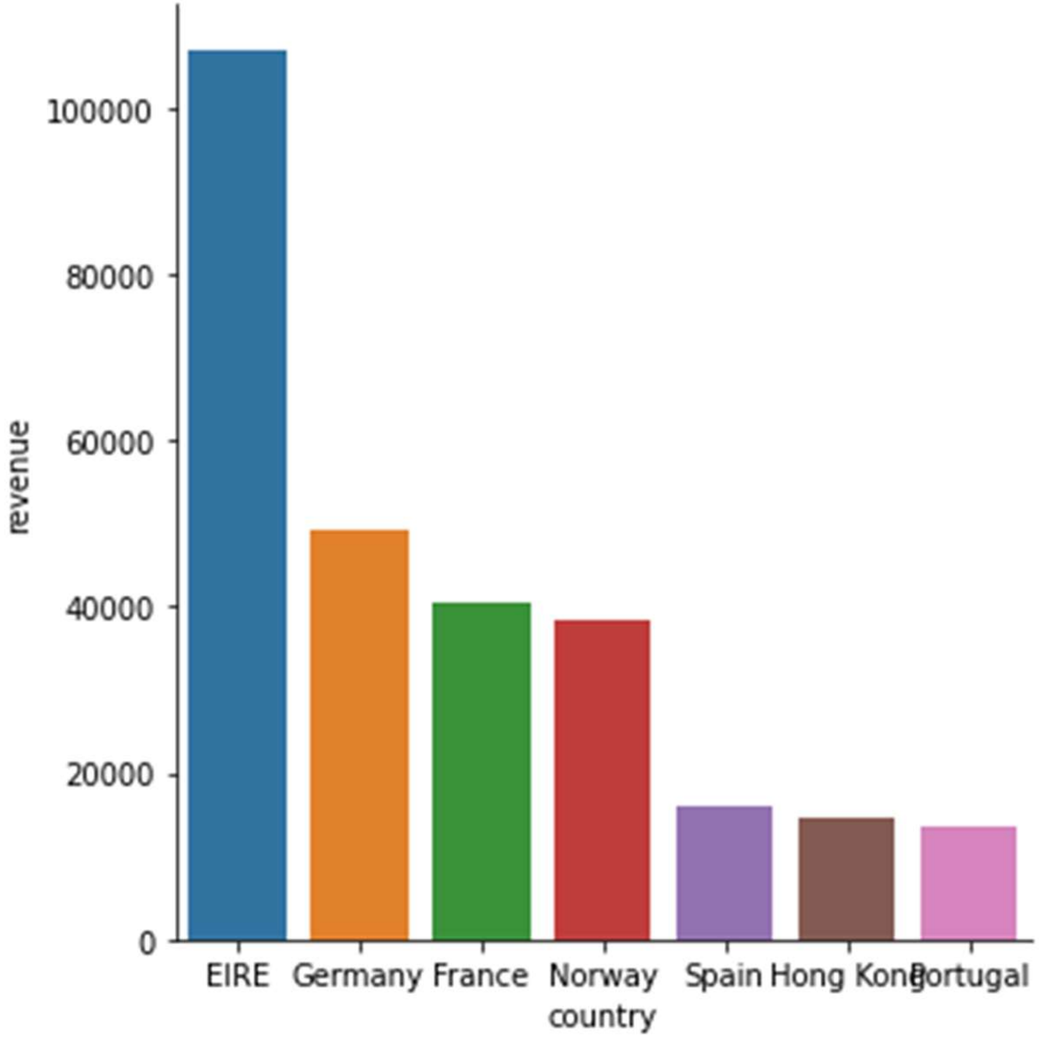
5. Articulate your findings using a deliverable with visualizations.

- Summarize what you have learned in your investigations using visualizations.
- (THIS PRESENTATION)

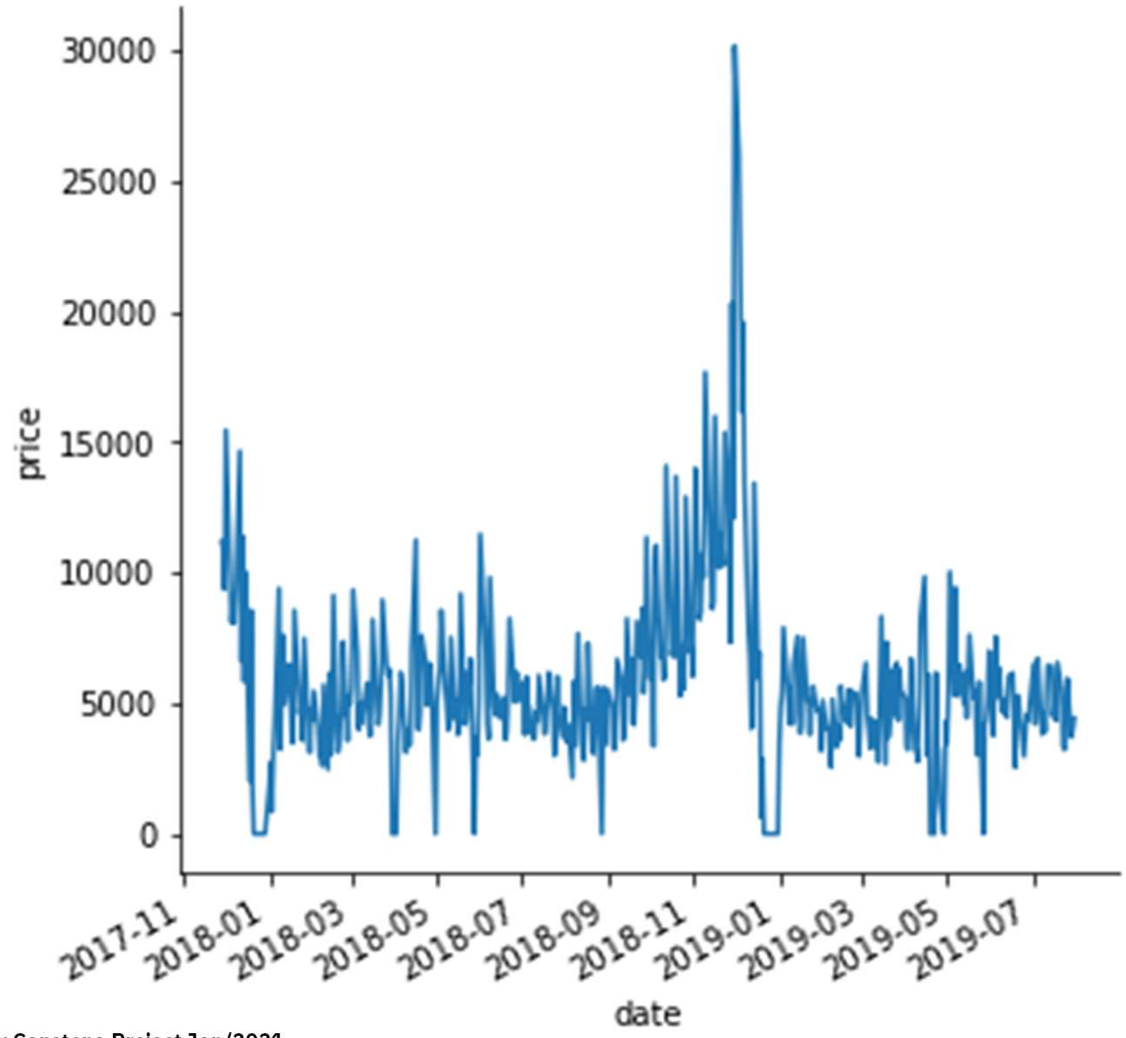
Revenue by Country (First 5 with UK)



Revenue by Country (First 7 without UK)



Revenue (price) by date (Peak on Nov/Dec)



Analytic Approach (Phase 2) - TBD

Express problem in context of statistical and machine learning techniques

- Machine Learning in Python
- Simple and efficient tools for data mining and data analysis

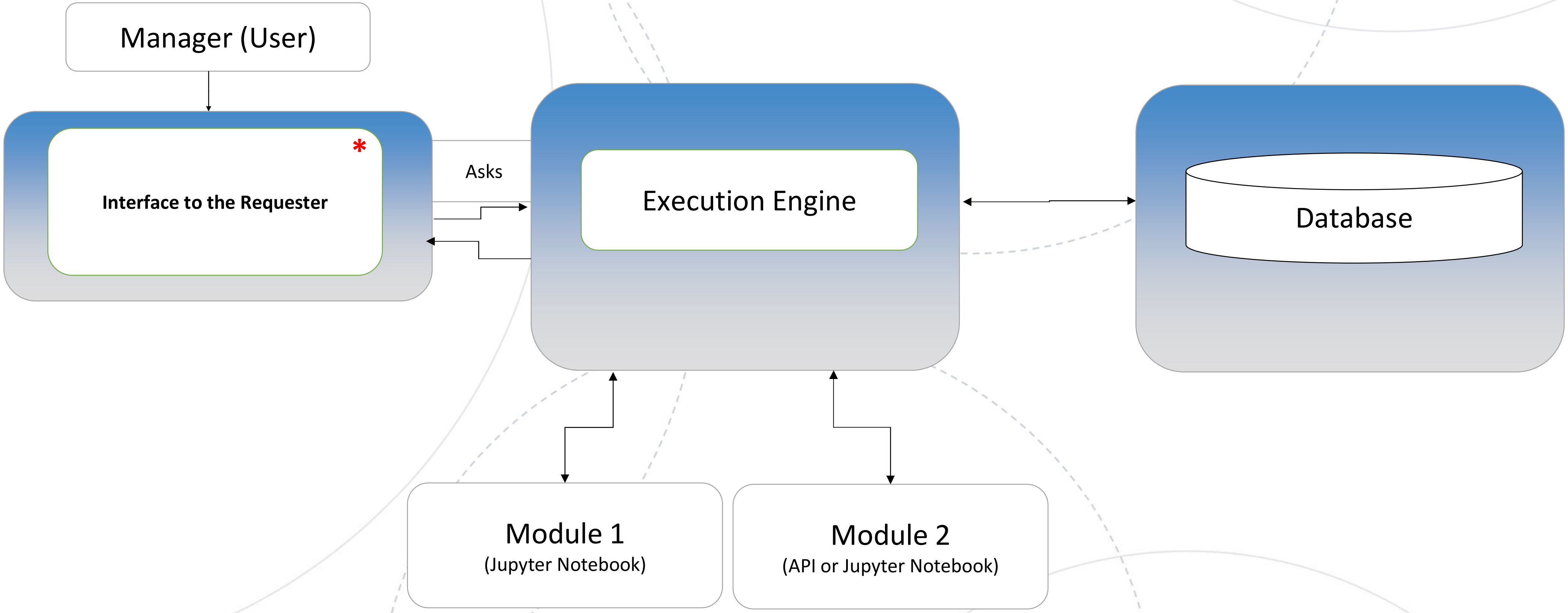
The analytic approach will consist of:

- TBD

View of Sprint and Phases

	Sprint 0 Phase 1	Sprint 1 Phase 2	Sprint 2 and beyond Phase 3	Sprint 3	
DEPLOYMENT SCENARIOS	EAD	NLP – ML Solution	Production	Improvements	**COMMENTS**
1. EU Countries		✓TBD	❑ Improve accuracy ❑ Improve spent time with forecast ❑ TBD	✓TBD	
2. Other GEO (Americas)			❑TBD	✓TBD	• TBD
3. TBD		✓TBD	✓TBD ❑TBD ✓TBD	✓TBD	

High Level Architecture (TBD Model)



*TBD

References

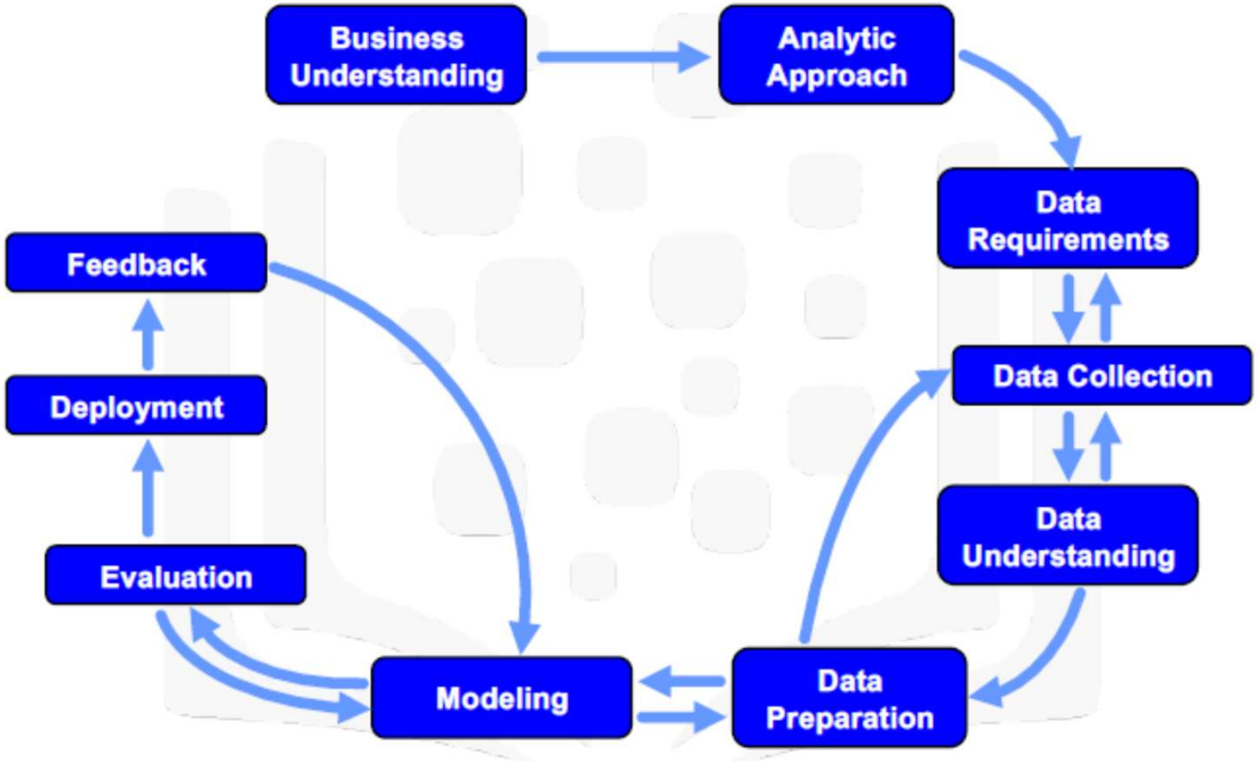
1. TBD

Back-ups

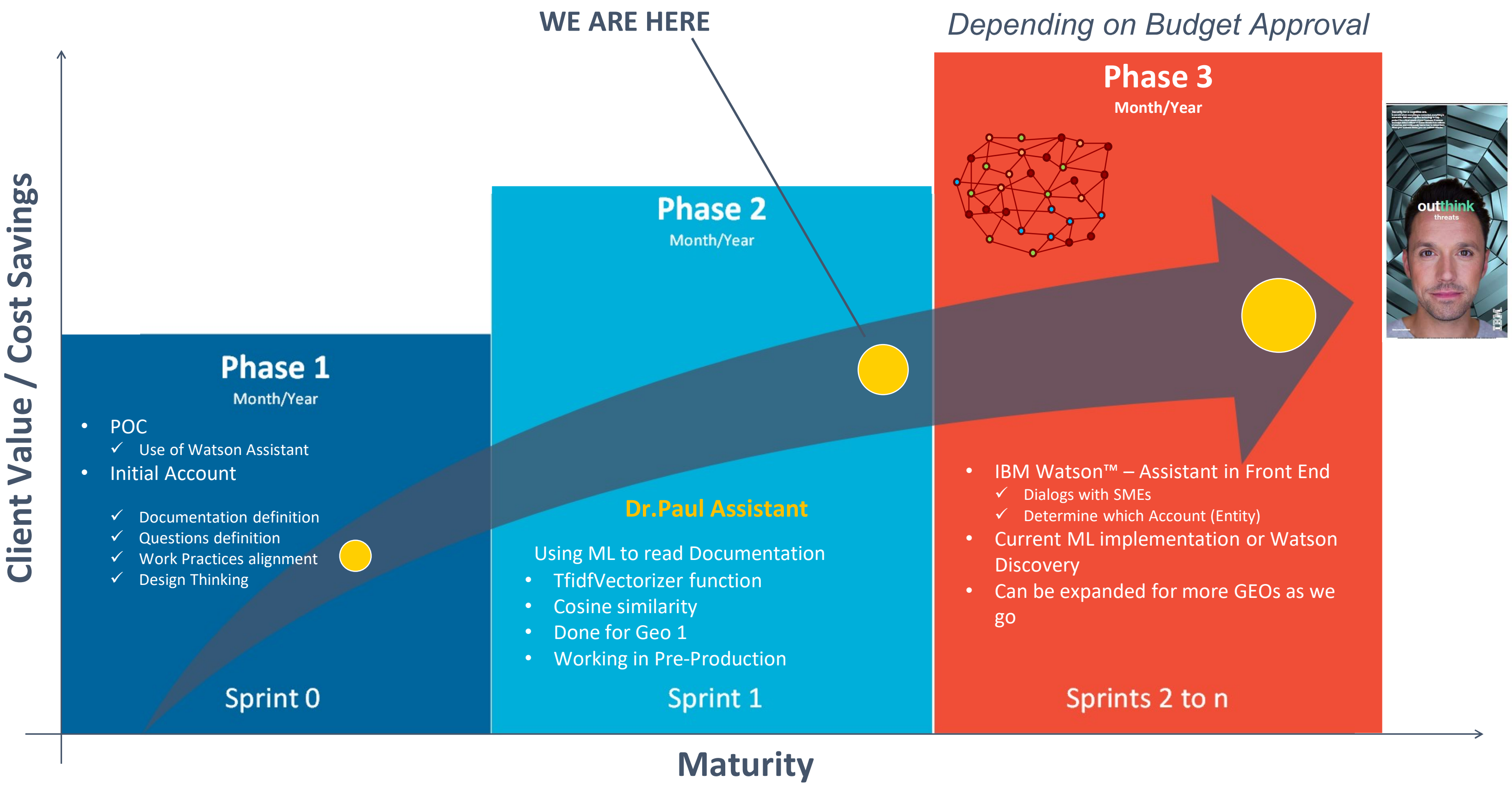
(Slides with some ideas or back-up info)

Methodology overview

CRISP-DM Methodology diagram



Phases approach (Example)



WE ARE HERE

Depending on Budget Approval

Phase 1

Month/Year

- POC
  - ✓ Use of Watson Assistant
- Initial Account
  - ✓ Documentation definition
  - ✓ Questions definition
  - ✓ Work Practices alignment
  - ✓ Design Thinking

Sprint 0

Phase 2

Month/Year

Dr.Paul Assistant

Using ML to read Documentation

- TfidfVectorizer function
- Cosine similarity
- Done for Geo 1
- Working in Pre-Production

Sprint 1

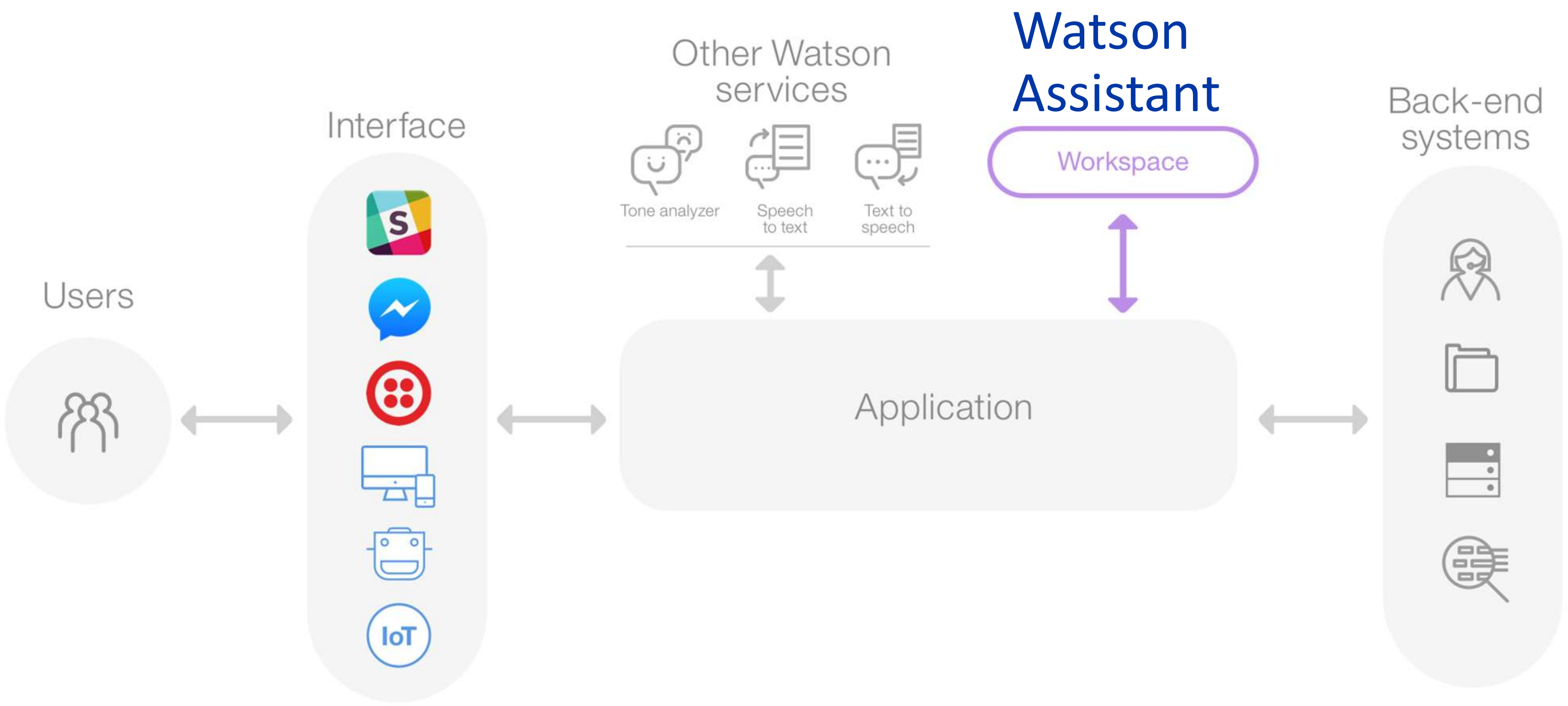
Phase 3

Month/Year

- IBM Watson™ – Assistant in Front End
  - ✓ Dialogs with SMEs
  - ✓ Determine which Account (Entity)
- Current ML implementation or Watson Discovery
- Can be expanded for more GEOs as we go

Sprints 2 to n

Architecture for Watson Assistant Solution (Typical – Example)



Users

Interface

Other Watson services

Tone analyzer

Speech to text

Text to speech

Watson Assistant

Workspace

Application

Back-end systems