# Roche A

Roche is a Swiss multinational healthcare company that is the world's largest biotech company. They are a leading provider of in vitro diagnostics, a global leader in cancer treatments and a frontrunner in personalized healthcare. This project is working with a Roche U.S. team.

**Project Description**

Roche engineers regularly onsite complete customer "service activities" onsite to perform both planned and unplanned maintenance on medical instruments. The objective of this case is to develop machine learning models to predict future service activities for any given customer/instrument/geography during a specific time interval (e.g. month, week, day). This will afford Roche the opportunity to better plan how/when service could be performed and better accommodate future staffing levels to meet the demands of future service activities.

Exploratory data analysis (EDA) will be conducted to identify and quantify the likelihood of predictive characteristics, attributes, and model focus areas. Unsupervised clustering will also be performed to identify additional patterns. Time series modeling will be leveraged to predict the likelihood of future service activities. The predicted likelihood of future service activity over time will be reported by customer and by instrument, and the predicted volume of future service activity will be reported based on geographic region.

**Internal Partners:** Justin Kundert - Data Science Manager at Roche Diagnostics

**Datasets**

- Service Activity History (date of service, type, instrument, customer, volume)
- Customer Data (ID, location, customer type)
- Instrument Data (instrument line/type/age/etc)

**Preferred Coding Languages:** Python

**Specific Skills**

1. Exploratory Data Analysis: Investigate raw dataset provided by Roche and prepare a report summarizing EDA findings and how they guide machine learning model development
2. Machine Learning Modeling: Using the provided dataset to develop models predicting service activity load for Roche technicians. Experience with unsupervised clustering and time series analysis is a plus.

Expected Technical Difficulty: **Intermediate/Advanced**

# Roche B

Roche is a Swiss multinational healthcare company that is the world's largest biotech company. They are a leading provider of in vitro diagnostics, a global leader in cancer treatments and a frontrunner in personalized healthcare.

## Project Description

Roche conducts a plethora of studies and administers several treatments to patients across the world. They would like the HDAG team to develop a Patient and Society Benefit Metrics (PBSM), leveraging in particular the initial work conducted by the PBSI team (Patient Benefit and Societal Impact). The PBSI team has created an informative framework, defining key "domains" where benefits can be understood from a "Patient" and/or "Society" perspective. However, these are not quantified or measurable, and cannot be used for insights or comparison amongst Roche's portfolio of studies.

The team will work to develop a concise set of the most informative metrics on Patient and Society benefit and validate it both internally and externally. This will ultimately inform study design decision making through insights about the benefits that specific study designs have upon patients and society.

This work will contribute towards the development of trustworthy scientific evidence that will extend the benefit of therapies, services, and integrated solutions to broaden access to more patients and accelerate development. Aligned to a 1-3 year strategic horizon, Roche would like to accelerate the development and validation of PBSM for ultimate adoption and operationalisation at the Disease Area (DA) level.

## Internal Partners:

- Ben Geelan (Strategic Operational Intelligence)
- Ilias Charlafti (Community Health Engagement Director)
- Detlef Hold (Head of Digital Learning Experience at Roche)

**Datasets:** Roche-provided dataset once the MVP is outlined

**Preferred Coding Languages:** Python, R (Shiny)

## Specific Skills

1. Data Consulting: Determine important and relevant data sources to analyze in the future
2. Data Analytics: Exploring, processing and deriving valuable insights from data
3. Data Visualization: Creating useful and interactive visualizations

Expected Technical Difficulty: **Intermediate**

# Roche C

Roche is a Swiss multinational healthcare company that is the world's largest biotech company. They are a leading provider of in vitro diagnostics, a global leader in cancer treatments and a frontrunner in personalized healthcare.

**Project Description**

As a pharmaceutical company, Roche requires government approval to bring its products to market. Each of these filings requires significant documentation: Roche estimates that each submission requires around 1500 documents (~2000 pages) of text. There is significant overlap between separate documents: often, all that needs to be changed is a year or molecule. Roche wants a system to automatically find documents similar to the one they are trying to produce to accelerate the filing creation process. **This project is a continuation of HDAG's engagement from last semester where a case team built an NLP model framework** to determine, assess, and visualize a similarity network of content.

Determining reusability manually would be difficult: at roughly 45 2000-page filings a year, Roche estimates that it would take 25 years to manually standardize documents. As such, Roche would like to develop a model to compute document reusability paths and establish an optimization approach for multi-reusability paths.

The project has three parts. First, the team will create a framework for network visualization based on the document similarity metrics developed last semester. Second, the team will develop an algorithm for reusability path computation between different documents in the network. Finally, the team will develop and refine an optimization approach for multi-reusability paths.

**Internal partners**: This project is in collaboration with the Strategic Operational Intelligence team at Roche.

**Dataset**: Roche intends to give us access to historical filings from past clinical trials, starting with the past twenty years of documentation for 50 molecules.

**Preferred coding language**: Python

**Specific Skills**: Network visualization, machine learning, neural network-driven NLP, discrete optimization

Expected Technical Difficulty: **Advanced**