

# Exploratory data exercise

---

## Contents

Objective .....	1
Business background .....	2
Your task .....	2
Output and deadline .....	3

## Objective

The objective of this exploratory data exercise is to assess:

1. relevant technical skills in data exploration, text mining, machine learning / forecasting and
2. fundamental business understanding.

This exploratory data exercise is not an offer for employment or to contract for the scope included. Issuance of this document, participants' preparation, participation in meetings and presentation of results and subsequent receipt and evaluation of those results by Sandoz does not commit Sandoz to award any sort of contract to any participant. Sandoz will not be liable for or reimburse any expenses incurred by the participants in relation to the participation in this data exercise. Only the complete execution of a written contract by both parties will obligate Sandoz in accordance with the terms and conditions contained in such a contract.

Participants have the right to opt out of the data exercise at any moment without providing any reasons. Sandoz reserves the right to select participants and / or exclude participants and / or change the rules and / or the length of this data exercise without providing any reasons.

By participating, participants agree to the rules of the exploratory data exercise.

This document is Sandoz Confidential Information which is in addition copyrighted material which may not be copied, recreated, or reused in any way or format.

## Business background

The European Patent Office (EPO) is the central institution responsible for granting European patents under the European Patent Convention (EPC). It examines and grants European patents, providing protection in up to 44 European countries. Then, a single application can lead to patent protection across multiple jurisdictions.

The EPO is crucial for understanding the patent landscape in Europe. By accessing EPO databases, companies can identify existing patents and avoid potential infringements. In particular, Generic pharma companies need to ensure that their products do not infringe on active patents. Generics companies often challenge the validity of existing patents to clear the way for their products. The EPO's opposition procedure allows third parties to challenge the validity of a granted patents within nine months of its publication.

Understanding patent expiration dates is critical for generics companies to time their market entry. The EPO offers detailed patent information, including expiry dates. After the basic patent expires, Supplementary Protection Certificates (SPC)s and pediatric extensions (PED)s can extend the protection of pharmaceutical products for up to 5,5 additional years. Generics companies must be aware of SPCs and PEDs when planning product launches.

Filing a single European patent application is often more cost-effective than filing multiple national applications, which can be beneficial minimizing legal costs. EPO allows to engage in opposition proceedings to challenge competitors' patents when necessary. The Boards of Appeals (BoA) is the final judicial instance in the proceedings: <https://www.epo.org/en/case-law-appeals>

Where appeals resolutions can be searched online (using Application Number, for example):

<https://register.epo.org/regviewer>

As you can imagine, Intellectual Property departments are interested in keeping an eye on the EPO's decisions and new patent filings relevant to their therapeutic areas.

## Your task

Help the IP team to better understand patent data leveraging your Data Science skills. You will develop a small Proof of Concept on European patents for pharma industry.

You will be given a database of collected pharma related BoA documents. Per document, you will have key fields such as:

- Decision date: when the decision is made public
- Case number
- Application number
- Publication number

- IPC pharma: whether International Patent Classification code is on A61K for pharma drugs
- IPC biosimilar: whether code related to B12N (biosimilars)
- IPCs: all IPCs related to the document
- Language
- Title of Invention
- Patent Proprietor
- Headword:
- Provisions:
- Keywords:
- Decisions cited: set of prior BoA documents used as a reference for current appeal
- Summary: Summary of Facts and Submissions
- Decision reasons: set of reasons to support the appeal decision
- Order: final appeal decision
- Order status: Order decision takeaway
- Order status web: Order decision takeaway extracted from website
- Order status manual: Order decision takeaway mapped manually
- Opponents: short list of companies presenting the appeal
- Opponents\_rep: pairs of individual opponents participating in the appeal against the Patent Proprietor with their respective law firm and patent holder law firm
- Text: consolidated version in a single body of prior fields
- xlm\_backup: data recovery backup of all [boa available fields](#)

The documents will be provided as:

- Json file: list of documents, each document as dictionary
- Excel file: main fields formatted as table

Given the documents, build a python-based Proof of Value that covers:

- ✓ Exploratory Data Analysis on the document's repository and key insights on appeals decisions (patent revoked, patent granted, patent maintained as amended, appeal dismissed, etc)
- ✓ Predictive approach on patent revocation likelihood
- ✓ Approach presentation to business stakeholders

Out of scope but positive valuation:

- ✓ Web-scrap additional / missing information if any
- ✓ LLM chatbot to interact with the pool of documents

## Output and deadline

**Please return by email by the end of the time allowed:**

- Presentation (ppt)
- Reproducible code (repository, .py, notebooks)

You will be asked to present the results as well as the followed approach, assumptions, and future improvements.

**Thank you for participating in the data exercise. We wish you best of luck!**