

---

# Nuruddin Bahar

Seventeen projects spanning  
bioprocess engineering, biomaterials  
and DNA nanotechnology to production  
data analytics and AI.

**17**

PROJECTS

**4**

CAREER ERAS

**173**

COURSES

[nbahar@musingsinbiology.com](mailto:nbahar@musingsinbiology.com)

# Contents

|    |  |                              |
|----|--|------------------------------|
| 00 | Career Trajectory  | TIMELINE                     |
| 00 | About — Lifelong Learning                                  | PROFILE                      |
| 01 | Healthchecks & Report TAT Dashboard                        | POWER BI · ONEHUB            |
| 02 | Apollo HealthBridge — Clinic Performance Dashboard         | POWER BI · SEMANTIC MODEL    |
| 03 | Apollo Diagnostics — Network Performance Dashboard         | POWER BI · DAX               |
| 04 | Apollo QA Chatbot — Conversational Analytics Assistant     | FASTAPI · RAG · LLM          |
| 05 | Apollo Clinics Ranking — Network Performance & Leaderboard | POWER BI · PBIP/TMDL         |
| 06 | Insta Data Automation                                      | PYTHON · ETL                 |
| 07 | VBA Robotic Batch Reporting Application                    | EXCEL VBA · AUTOMATION       |
| 08 | Statistical Analysis for Process Development               | STATISTICS · BIOPROCESS      |
| 09 | Impurity Analysis of Analytical Data                       | PYTHON · POWER BI · QC       |
| 10 | Design of Experiments (DOE) GUI                            | PYTHON · TKINTER             |
| 11 | Python Resume Maker  | PYTHON · MATPLOTLIB          |
| 12 | Django Website Platform                                    | FULL-STACK · DJANGO          |
| 13 | Bioprocess Risk Assessment Tool                            | EXCEL VBA · FMEA             |
| 14 | Golden Batch Analysis                                      | PYTHON · MULTIVARIATE        |
| 15 | DNA Origami Magnetic Nanopropellers                        | M.TECH THESIS                |
| 16 | Mechanoresponsive PEG-Laminin Hydrogels                    | MSC · BIOMATERIALS           |
| 17 | Scientific Review Papers                                   | TECHNICAL WRITING · RESEARCH |

# Career Trajectory

Two decades from school and engineering to biotechnology research and production data analytics — milestones grouped by era and coded by category.

● Academics ● Profession ● Certification

## 2006 – 2014

- **ACADEMICS · SCHOOL**  
Secured 93.4% in CBSE 10th with 100% in Social Science
- **ACADEMICS · SCHOOL**  
Secured 91% in CBSE 12th
- **ACADEMICS · EXAM**  
Secured All-India Rank 6006 in IITJEE 2008
- **ACADEMICS · EXAM**  
Secured All-India Rank 718 in IITJEE 2009
- **ACADEMICS · GRADUATION**  
Started B.Tech (Chemical Engineering) at IIT Bombay
- **PROFESSION**  
Internship at Nth Wall Entertainments Pvt. Ltd. in Mumbai
- **PROFESSION**  
Internship at Tridiagonal Solutions Pvt. Ltd. in Pune
- **ACADEMICS · GRADUATION**  
Graduated from IIT Bombay after completing B.Tech in Chemical Engineering
- **ACADEMICS · EXAM**  
Cleared CSIR-JRF (Life Sciences) on first attempt with All-India Rank 42
- **PROFESSION · SOFTWARE**  
Worked as Associate Software Engineer at Exeter Software Pvt. Ltd., Bengaluru

## 2014 – 2019

- **ACADEMICS · POST-GRADUATION**  
Started M.Tech (Nano Science and Engineering) from IISc Bengaluru

ACADEMICS · EXAM

Cleared CSIR-JRF (Life Sciences) on second attempt with All-India Rank 81

ACADEMICS · EXAM

Cleared GATE (Biotechnology) by securing All-India Rank 16

ACADEMICS

Presented poster at Chemical Engineering Symposium at IISc Bengaluru

ACADEMICS

Presented poster at International Conference on Emerging Electronics

ACADEMICS · POST-GRADUATION

Completed M.Tech thesis 'Towards DNA Origami Based Magnetic Nanopropellers' under Prof. Ambarish Ghosh and Prof. Banani Chakraborty

ACADEMICS · GRADUATION

Completed M.Tech with CGPA 5.9 and grade A in thesis

ACADEMICS · BIOTECHNOLOGY

Completed advanced course on 'Cell culture, cell line derivation, and human disease biology' from CCMB Hyderabad

ACADEMICS · POST-GRADUATION

Started MSc (Biomedical Engineering) at University of Glasgow

ACADEMICS · POST-GRADUATION

Completed thesis work on 'Design and Characterization of PEG-Laminin based Mechano-responsive Hydrogels for Drug Delivery and Gene Therapy'

2019 – 2023

ACADEMICS · GRADUATION

Completed MSc (Biomedical Engineering) at University of Glasgow with Distinction

PROFESSION · BIOTECHNOLOGY

Worked in downstream processing as Research Scientist at Wockhardt Research Centre, Aurangabad, INDIA

PROFESSION · BIOTECHNOLOGY

Worked in upstream processing as Research Scientist at Wockhardt Research Centre, Aurangabad, INDIA

PROFESSION · BIOTECHNOLOGY

Working in bioprocess engineering and MSAT as Research Scientist at Wockhardt Research Centre, Aurangabad, INDIA



CERTIFICATION

Microsoft Word Expert



CERTIFICATION

Microsoft Excel Expert



CERTIFICATION

Microsoft PowerPoint Associate



CERTIFICATION

Microsoft Outlook Associate



CERTIFICATION · SOFTWARE

Certified Associate in Python Programming (Python Institute)

2023 – Present

---



PROFESSION

Joined Dr. Reddy's Laboratories as a Data Analyst



CERTIFICATION

Microsoft Certified: Data Analyst Associate (PL-300)



CERTIFICATION

AI Practitioner Certification



PROFESSION

Joined Apollo Health and Lifestyle Ltd. as a Data Analyst

# A Lifelong Learner

173

LINKEDIN LEARNING  
COURSES COMPLETED

Software · MS Office ·  
Python · Django · Adobe ·  
professional skills

I have a strong belief that learning should be a lifelong journey, and I consistently strive to apply transferable skills to my daily work. To fuel my continuous learning, I have chosen to utilize LinkedIn Learning and its wide range of courses. Over the years, this platform has significantly enhanced my knowledge and skill set, contributing to my professional development.

I have successfully completed numerous LinkedIn Learning courses, covering diverse areas such as software skills, MS Office Suite, Python, Django, Adobe applications, and personal and professional skill development. Although some courses have been retired, I have a total of 173 courses under my belt. These courses have been instrumental in expanding my expertise and enabling me to tackle various challenges in my professional endeavors.



# Healthchecks & Report TAT Dashboard

---

## ABSTRACT

The Healthchecks & Report TAT Dashboard is a Power BI solution that gives clinic managers a single, real-time view of health-check footfall and diagnostic report turnaround across the ONEHUB network. It consolidates appointment, billing, and test-level data into one analytical model, making it immediately clear where reports stand and how long each test takes to travel from sample collection to final delivery.

## THE CHALLENGE

The operational data was scattered across multiple ONEHUB PostgreSQL tables, and the timestamps that matter most — when a sample was collected, transferred, started, and completed — were locked inside EPOCH values and nested JSON rather than queryable datetime fields. No one had a consolidated, cross-clinic view of bookings versus walk-ins, where reports were stalling, or which tests dragged turnaround. Managers assembled that picture by hand, one clinic at a time.

## DATA CONSOLIDATION & ENGINEERING

I authored a consolidating SQL query that joins the relevant tables through their key identifiers — mapping each clinic to its patients via UHID and Visit ID, and on to agreements, packages, and inclusions (the individual tests). EPOCH timestamps were converted to readable datetimes, and the precise collection, transfer, start, and end times were parsed out of JSON to reconstruct each test's true turnaround journey.

## DAILY PIPELINE

A Python pipeline extracts this data daily and stores it as master and clinic-specific CSV and Parquet files — a booked-appointments master, a billed-appointments master carrying the report status of every test, and a consultations table held in reserve. Power BI consumes these as a refreshable model behind the report.

## OPERATIONAL INSIGHTS & IMPACT

The test-level view ranks the busiest of 168 tests by their ready-versus-in-progress split and decomposes each test's turnaround into Collection, Processing, and Transfer hours, surfacing the slowest tests automatically. Managers read bookings, walk-ins, and footfall for a clinic or group across any day, week, month, or custom range — replacing manual, clinic-by-clinic tracking with one self-service view.

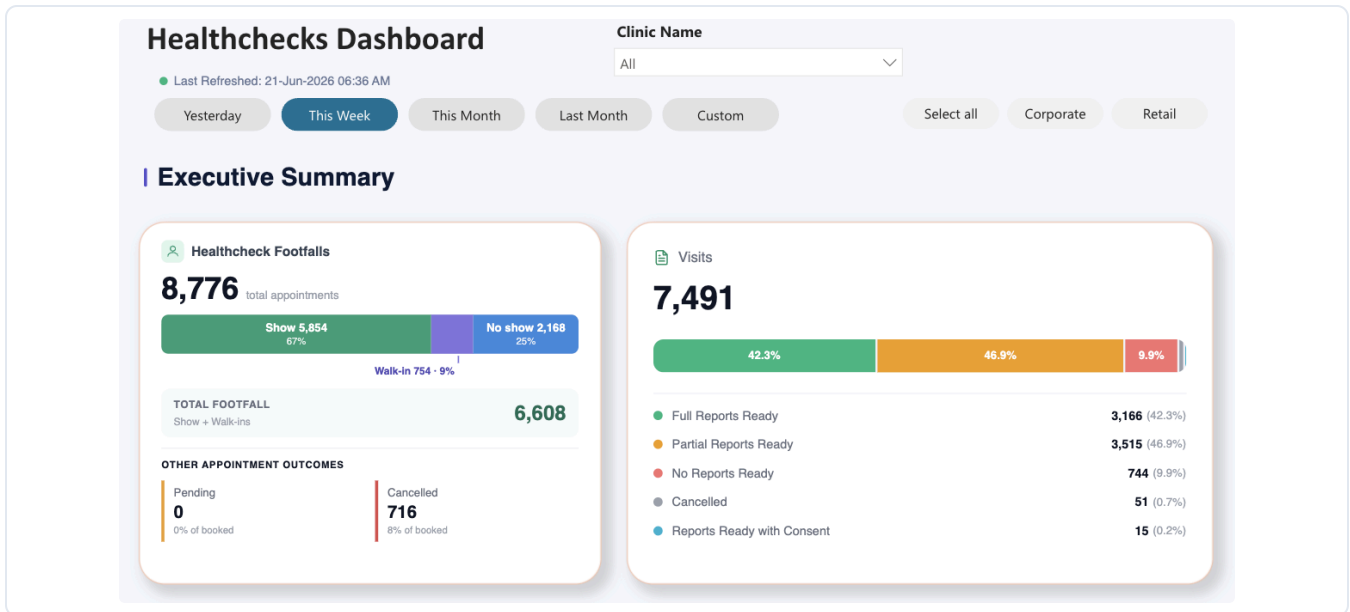


FIG. 01 Executive Summary — footfall, appointment outcomes, and visit-wise report readiness



FIG. 02 Operational Insights — test-wise report status and per-test TAT breakdown (Collection / Processing / Transfer)



# Apollo HealthBridge — Clinic Performance Dashboard

---

## ABSTRACT

The Apollo HealthBridge Dashboard is a Power BI reporting suite that gives every Apollo clinic a unified, self-service view of its business performance — revenue, footfall, patient loyalty, appointments, and referral conversion — all in one place. It is published in two forms: a per-clinic template each manager uses for their own unit, and an all-clinics super-admin template that lets leadership read overall averages and compare performance clinic by clinic.

## THE APPROACH

Rather than rebuild a report for every clinic, I designed HealthBridge as a thin, live-connected report sitting on a single published semantic model. The data and all the DAX logic live centrally; each published template is a lightweight copy bound to that model, keeping metrics perfectly consistent across clinics and letting updates propagate from one place. The model is refreshed by the same daily Python pipeline that powers the wider ONEHUB platform.

## TIME-GRAIN DESIGN

Users needed to read the same story at different resolutions, so I built a Day / Week / Month / Year toggle driven by bookmarks — selecting a grain shows the matching visuals and hides the rest. Each grain is backed by its own measure family, with a consolidated layer handling trailing 'last-three-period' comparisons, and every measure isolates the Corporate and Retail segments.

## PAGE 1 — REVENUE & FOOTFALLS

Four headline KPIs frame patient quality up front — Corporate Repeat Rate, Retail Repeat Rate, Corporate-to-Retail Conversion, and New Patients — followed by Total Footfalls and Total Revenue trends, each decomposed by Business Segment and by Service Category (Doctor Consult, Health Check, Other Services).

## PAGE 2 — APPOINTMENTS & REFERRALS

The operational view follows the patient journey — an appointment funnel from booked through show / no-show to consultation — alongside Top 5 Corporates by volume and revenue, Top 5 Services, Top 5 Doctors, Referrals & Conversions by clinic type, and Appointments by Source with their turnout.

## IMPACT

It makes revenue, footfall, and repeat rates clearly visible at both levels — each manager sees their own clinic in detail, while leadership reads network-wide averages and clinic-wise comparisons — replacing scattered, inconsistent reporting with one trusted source.

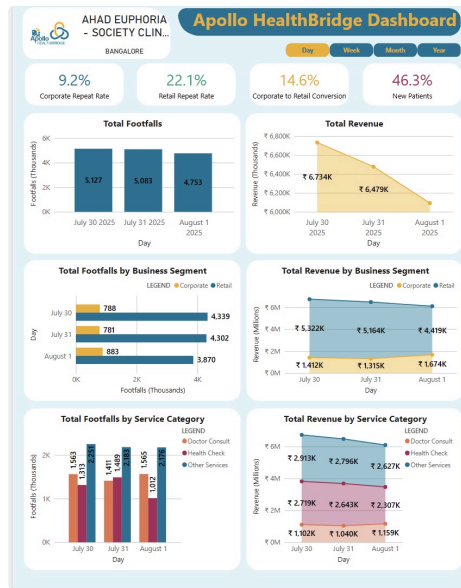


FIG. 01 Revenue & Footfalls — loyalty KPIs with footfall and revenue by business segment and service category

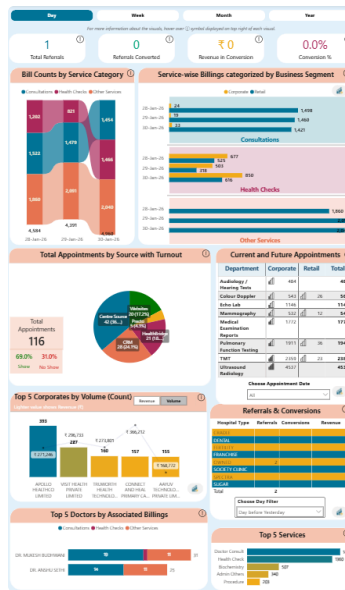


FIG. 02 Appointments & Referrals — funnel, top corporates / services / doctors, and referral conversion



# Apollo Diagnostics — Network Performance Dashboard

## ABSTRACT

A metric-driven Power BI dashboard for the Apollo Diagnostics network where one selector — Active Centres, Volume, Revenue, or AOV — re-themes the entire canvas, recalculating every visual, statistic, goal, and narrative to the chosen metric. It pairs Day / Week / Month / Year grain and multi-level slicers with KPI cards, a statistical goal engine, and a self-writing 'Summary based on Metric Goals' panel that turns the day's numbers into plain English. One report replaces four, with leadership able to pivot metric, scope, and period instantly.

## THE APPROACH

The report is driven by a small set of disconnected 'selector' tables — one for the active metric, one for the time grain (Day / Week / Month / Year), and parameter tables for Top-N and Last-N-Days. A central SWITCH layer reads the chosen metric and routes every measure — the KPI cards, the by-centre-type comparison, the daily trend, and the Top-N visuals — to the matching calculation, while the slicer panel (Centre Type, Centre Name, City, State, Business Zone) filters the underlying model. Changing the metric once makes the titles, colours, axes, and numbers all follow, so the report always stays internally consistent.

## METRIC-DRIVEN THEMING

Each metric carries its own colour identity — AOV in amber, Volume in green, Active Centres in blue — so the active view reads at a glance. The by-centre-type chart compares the current day against the previous one with signed % labels, and a 'Current Selection' banner echoes the live filter context so the scope behind the numbers is always clear.

## GOALS & SMART SUMMARY

A goal engine benchmarks each metric against a statistical target — Mode, Median, Mean, Max, or a chosen percentile — or a custom value, surfaced as headline goals. Beneath it, the 'Summary based on Metric Goals' panel writes itself: DAX-composed sentences report each metric against its goal and the previous day (e.g. '-96.6% vs. goal', '-78.5% vs. previous day'), turning the day's numbers into a plain-English read.

## INTERACTIVITY & TOP-N

Three parameterised panels dial the depth of the view — Last 2 / 5 / 10 / 30 / All days for the trend, and Top 2 / 5 / 10 / 20 / All for services and centres — so one report serves both a quick glance and a deeper drill. A Descriptive Statistics table profiles the selected metric's distribution (mode, mean, quartiles, max) across the network.

IMPACT

It gives the diagnostics network one daily control panel instead of four. A reviewer can pivot from AOV to volume, from the whole network down to a single centre or state, and read a written verdict on the day's performance against goal — without commissioning a new report.



FIG. 01 Network-wide view — AOV selected across all centres, with KPI cards, day-over-day comparison, statistics, goals, and Top-N panels

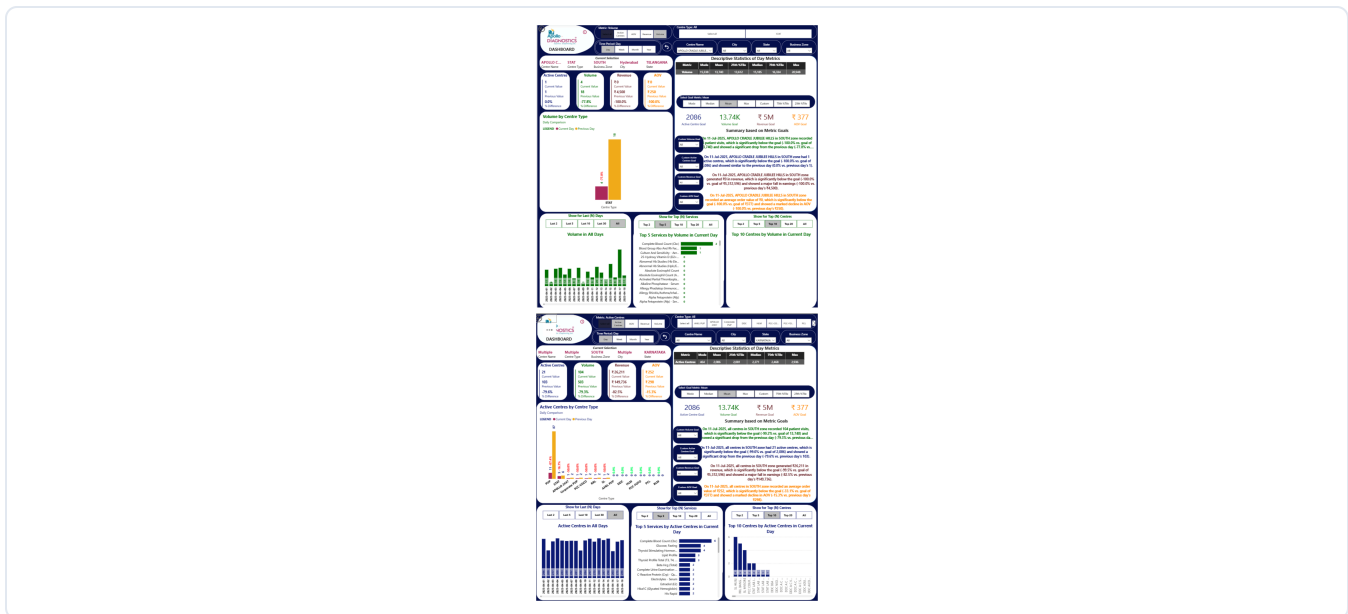


FIG. 02 Same canvas, different metric & scope — Volume on one STAT centre (green, top) and Active Centres across Karnataka (blue, bottom)



# Apollo QA Chatbot — Conversational Analytics Assistant

## ABSTRACT

A conversational analytics assistant that lets non-technical staff query Apollo's live databases in plain English. This FastAPI/Streamlit application converts a question into validated SQL, runs it, and returns a summarised answer with a downloadable table — grounded by a FAISS-based RAG layer over the real schema and foreign-key graph, routed through a deterministic cache for common roll-ups, and executed by a self-correcting agent that repairs failed SQL within strict, SELECT-only guardrails across two production databases. It collapses the gap between a business question and an answer from analyst-minutes to seconds.

## RAG & SCHEMA INTELLIGENCE

Table, column, and foreign-key 'cards' built from the schema catalogue are embedded with OpenAI's text-embedding model into a FAISS vector index. Each card carries a business-area tag (billing, facility, provider, corporate agreement) and sits in a foreign-key graph, so a question retrieves the right tables and joins instead of the model guessing column names — the single biggest factor in getting realistic SQL.

## HYBRID ROUTING & THE SQL AGENT

A deterministic router first parses the question for month, year, financial-year, quarter, and city, and serves matching totals straight from pre-computed aggregates — a fast path that skips the database entirely. On a miss, a bounded agent takes over: it generates SQL, validates it, executes it, and on any database error feeds the message back to the model to repair the query, retrying up to three times before summarising the result.

## SAFETY & GUARDRAILS

Generation is SELECT-only and capped to a row limit. A validator blocks cross-backend mistakes — MySQL syntax leaking into a SQL Server query and vice-versa — and checks every query against an allowed-table set. Credentials live only in environment variables and the database users are read-only, so the assistant stays close to live data without putting it at risk.

## IMPACT

It collapses the gap between a business question and an answer from minutes of analyst time to seconds of self-service, across two production databases, while a thread router and follow-up rewriter keep multi-turn conversations on topic. A feedback 'Learning Mode' that mines join patterns from good queries lets the schema knowledge improve with use.

BUILT WITH

- FastAPI
- uvicorn
- Streamlit
- FAISS
- OpenAI
- SQL Server
- MySQL
- SQLAlchemy
- pandas

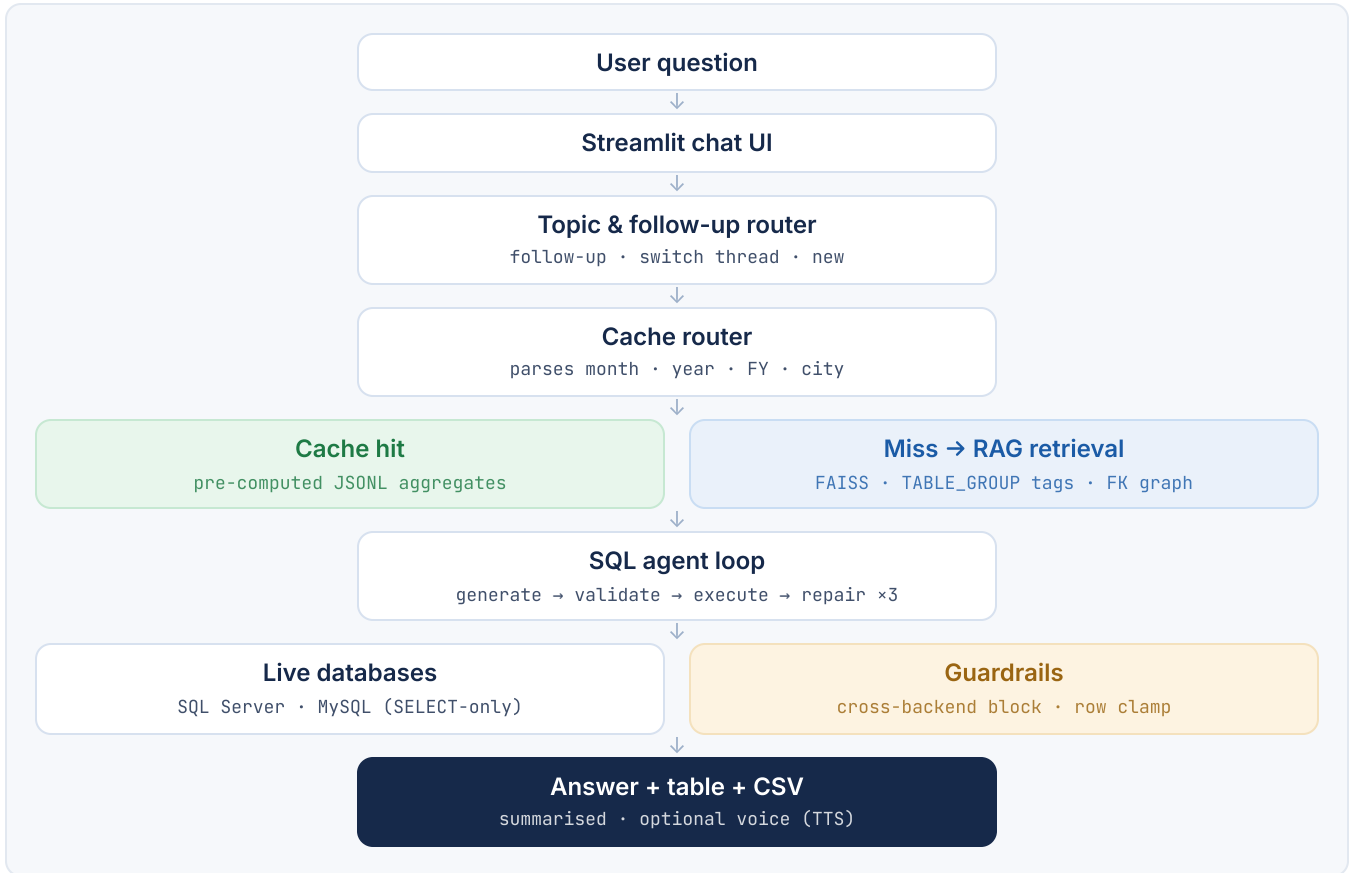


FIG. 01 End-to-end flow — a question is routed, grounded in the schema, turned into safe SQL, and returned as an answer

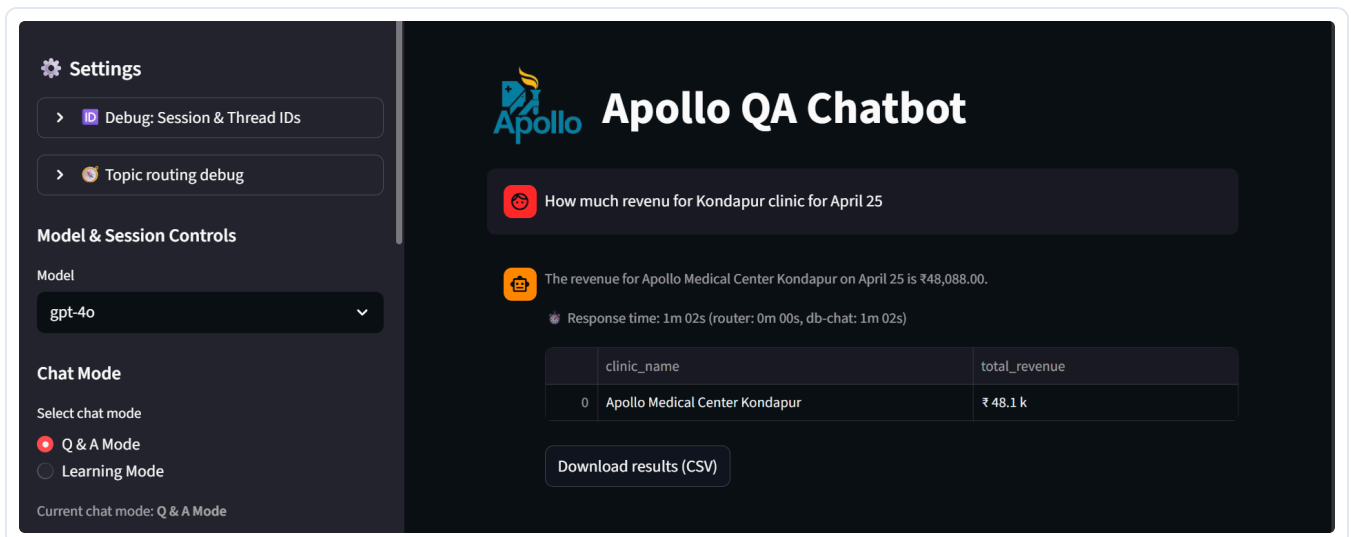


FIG. 02 The Q&A interface — a natural-language question resolved to live SQL, with a summarised answer and CSV export



# Apollo Clinics Ranking — Network Performance & Leaderboard

---

## ABSTRACT

A Power BI dashboard that ranks every clinic in the network against the others across revenue, visits, consultations, patients, doctors, efficiency, revisit rate, and discount leakage. Per-metric ranks combine into a single composite score for a definitive overall leaderboard, layered with dynamic performance tiers and a suite of HTML / DAX cards with inline SVG trendlines — all built in version-controlled PBIP/TMDL. It gives leadership one defensible answer to who to reward, who to coach, and where discounting erodes revenue.

## COMPOSITE RANKING ENGINE

Each clinic is ranked on revenue, visits, consults, patients, doctors, revenue per patient, revenue per doctor, revisit rate, and discount, with a small rank badge beside every figure. Summing those ranks gives one overall score, so the master table sorts to a definitive order — Anna Nagar, Kondapur, and Marathahalli surface at the top — while every column stays independently sortable for ad-hoc comparison.

## BEST / WORST LEADERBOARDS

A second view turns the same data into focused leaderboards — highest per-patient yield, highest revisit rate, top service departments, and highest visit volume, and on the other side the worst discount leakage and a watchlist of clinics needing attention. Each card names the leader, the network median, and the gap, turning a wall of numbers into a short, ranked story per metric.

## HTML DAX CARD SUITE

The headline tiles — network revenue, active clinics, network leader, and the visits / patients / doctors / consultations / revisit strip — are custom HTML DAX measures with inline SVG sparklines, so each shows its current value, year-on-year movement, and a trend at a glance, and re-renders instantly as the city, clinic, or financial-year slicers change.

## IMPACT

It replaces scattered clinic reports with one defensible ranking leadership can act on — who to reward, who to coach, and where discounting is eroding revenue. Because it is built in PBIP/TMDL, the whole model is version-controlled, so the ranking logic and tiers evolve through code review rather than ad-hoc edits.

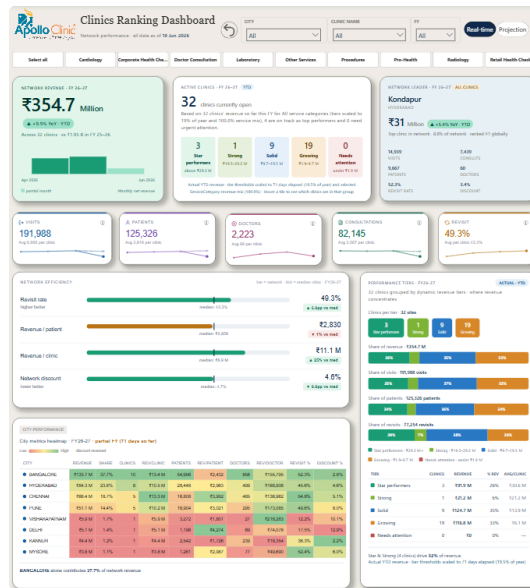


FIG. 01 Network overview — revenue, active clinics, the current leader, performance tiers, and a city-level metric heatmap

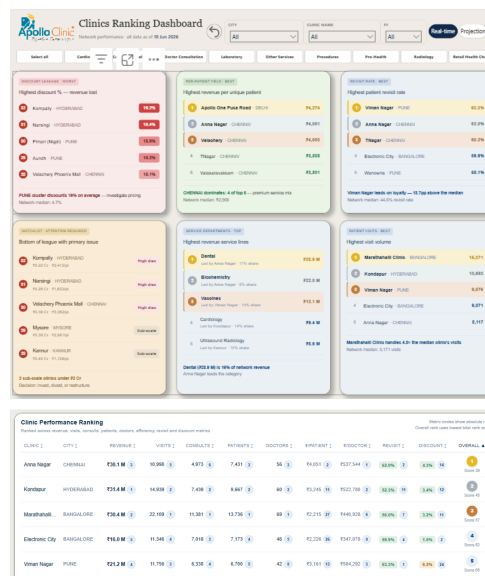


FIG. 02 Per-metric leaderboards (top) and the composite master ranking table with rank badges and an overall score (bottom)



# Insta Data Automation

## Daily Multi-Source ETL Pipeline

---

### ABSTRACT

Insta Data Automation is the data backbone behind the Apollo clinic dashboards — a Python ETL pipeline that runs unattended every night, pulls from three production databases, and refreshes the CSV and Parquet files that feed every clinic-wise and super-admin report. It turns scattered operational systems into a single, consistent, dashboard-ready dataset, delivered to a remote server through a data gateway with no manual intervention.

### THE APPROACH

The pipeline is built around incremental, watermark-driven loads: each source is queried only for what changed since the last run, merged into rolling 30-day masters and a legacy store, and written back atomically so a dashboard never reads a half-written file. A single nightly run extracts from SQL Server, MySQL, and PostgreSQL, applies the business logic the dashboards depend on, and fans the result out into per-clinic, per-corporate, and network-wide slices — each saved as both CSV and a compressed Parquet sidecar.

### MULTI-SOURCE EXTRACTION & INCREMENTAL LOADS

Three systems feed the pipeline — SQL Server (Insta billing), MySQL (health-checks, doctor consultations, and AHCN appointments), and PostgreSQL (the ONEHUB platform). Each is read incrementally from a stored watermark, with full-snapshot modes available for back-fills. New rows are upserted into 30-day rolling masters and a growing legacy store, and old Parquet generations are purged on a retention schedule.

### TRANSFORMATION & BUSINESS LOGIC

Raw operational data is reshaped into the exact grain the dashboards expect — visit-status labelling, consent flags, report-readiness states, and per-test turnaround (Collection / Processing / Transfer) are all computed here, replicating Power Query logic in pandas so the heavy lifting happens once, server-side, rather than inside every report.

### RELIABILITY & OBSERVABILITY

Every write is atomic with retries, and each dataset is emitted as CSV plus a Snappy-compressed Parquet sidecar for fast dashboard loads. A structured step-event log records the status, row counts, and byte sizes of every stage; the run can also email clinic-level KPI summaries and push artefacts to SharePoint — so a failure is visible immediately rather than discovered as a stale dashboard.

## OUTPUTS & IMPACT

One nightly run produces a consistent set of per-clinic, per-corporate, and super-admin files, served to the dashboards through a data gateway on a remote server. It removes manual data preparation entirely, guarantees every report reads the same numbers, and scales to the whole network from one codebase — the foundation the HealthBridge, Healthchecks / TAT, and Clinics Ranking dashboards all sit on.

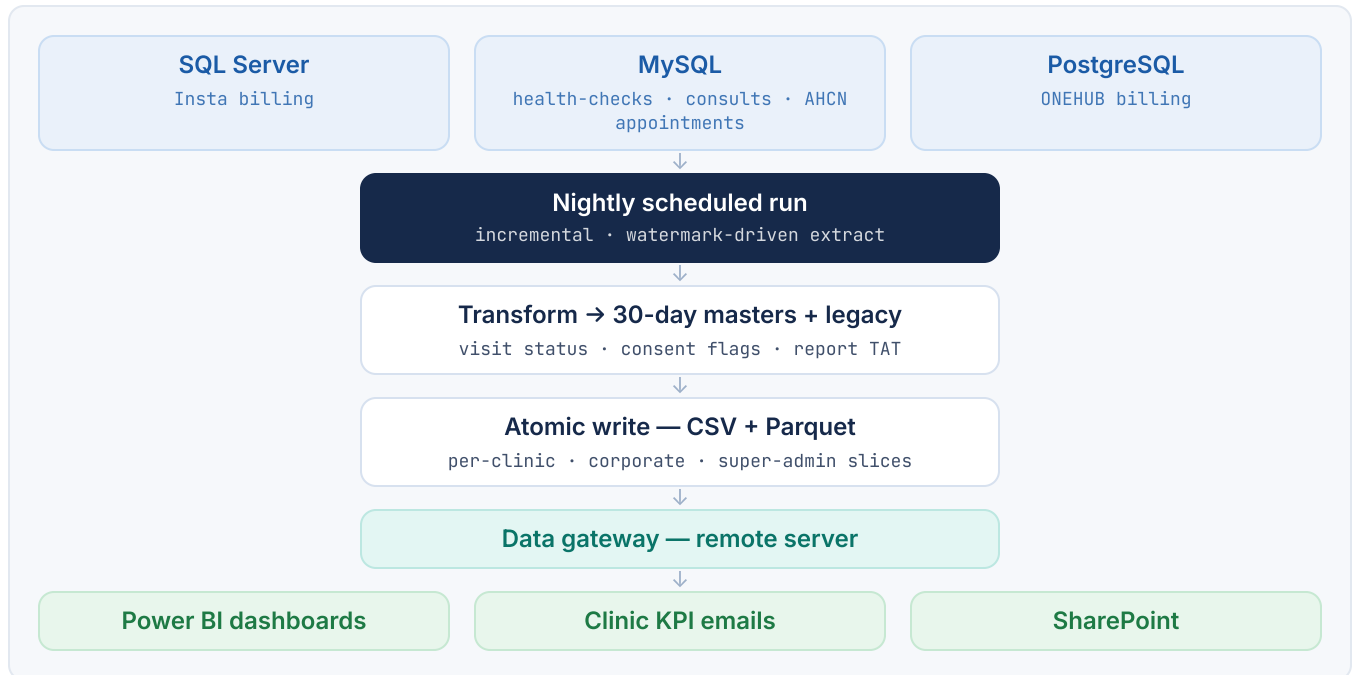


FIG. 01 End-to-end nightly flow — incremental extract from three databases, transform, atomic CSV / Parquet load, served via a gateway

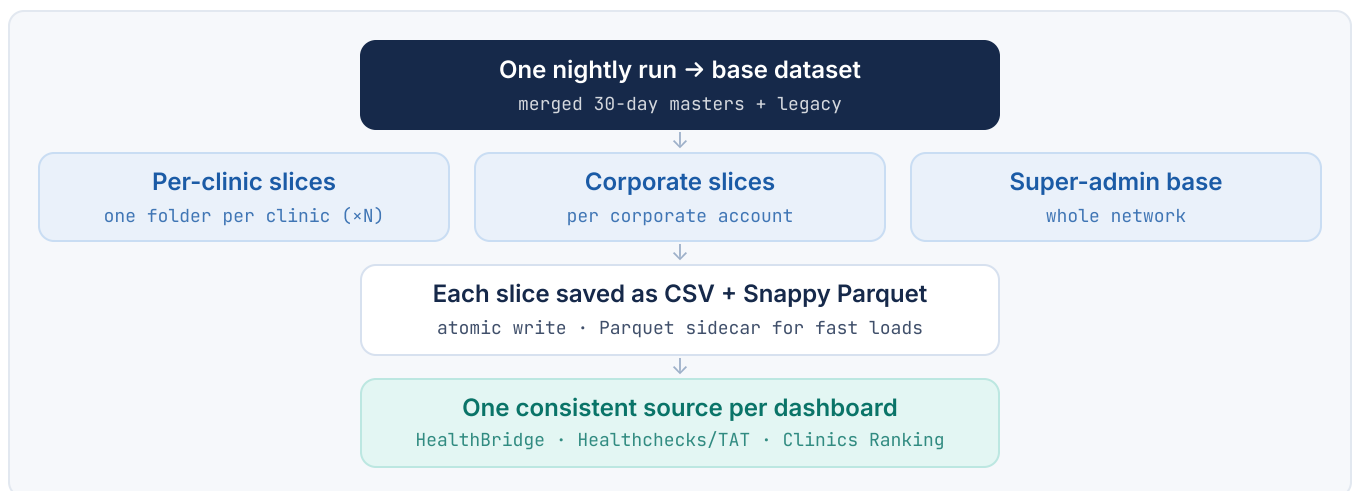


FIG. 02 Output fan-out — one run produces per-clinic, corporate, and super-admin slices that feed every dashboard



# VBA Robotic Batch Reporting Application

## OVERVIEW

Downstream-processing batch reports for insulin variants were repetitive and slow, so I automated the entire workflow in Excel VBA. I built a custom ribbon UI, linked UserForms, and modular routines — including 'Choose Product' and 'Change System' selectors — wired into a step-by-step wizard.

The wizard generates Word, Excel, and PowerPoint outputs on demand, turning a manual, error-prone chore into a one-click, standardised process. It was my first full VBA application.

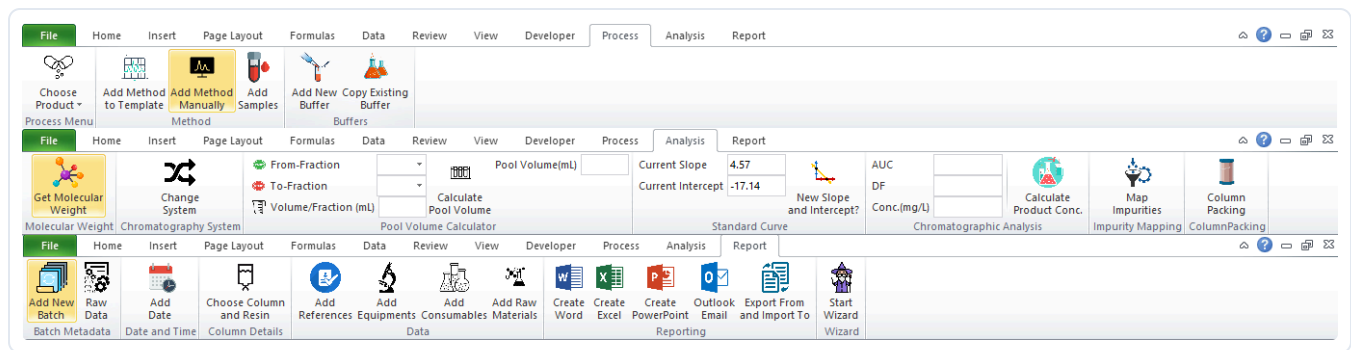


FIG. 01 Custom UI buttons (ribbon) created in Excel

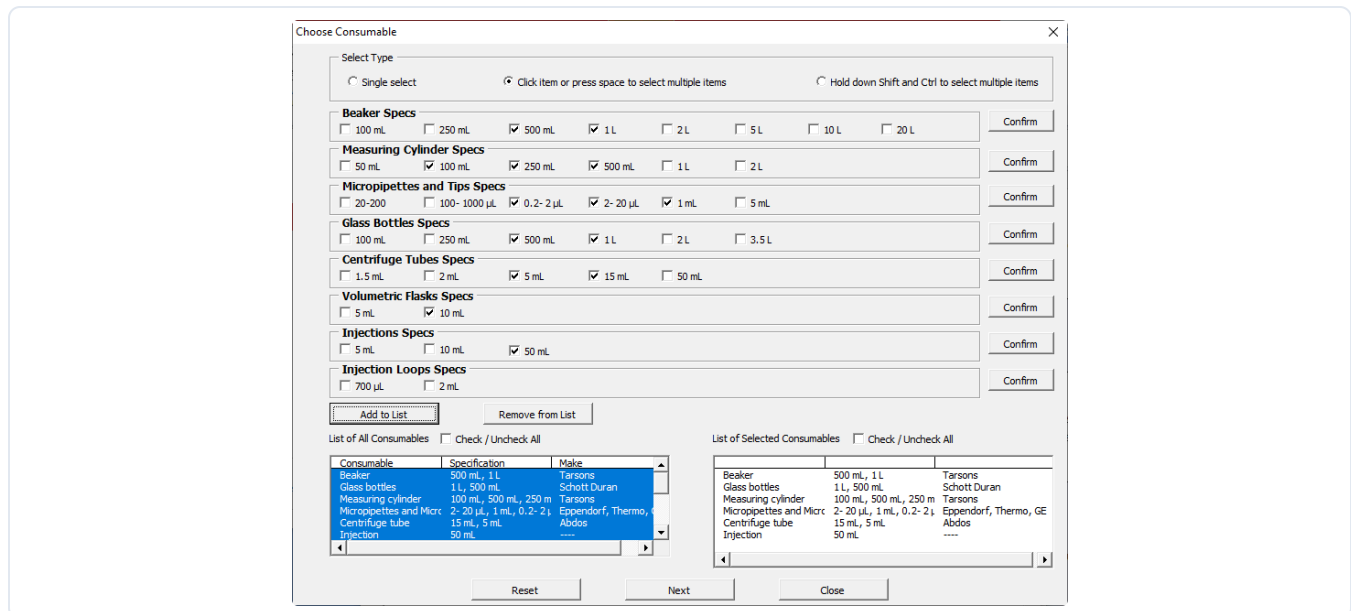


FIG. 02 A modular data-entry UserForm

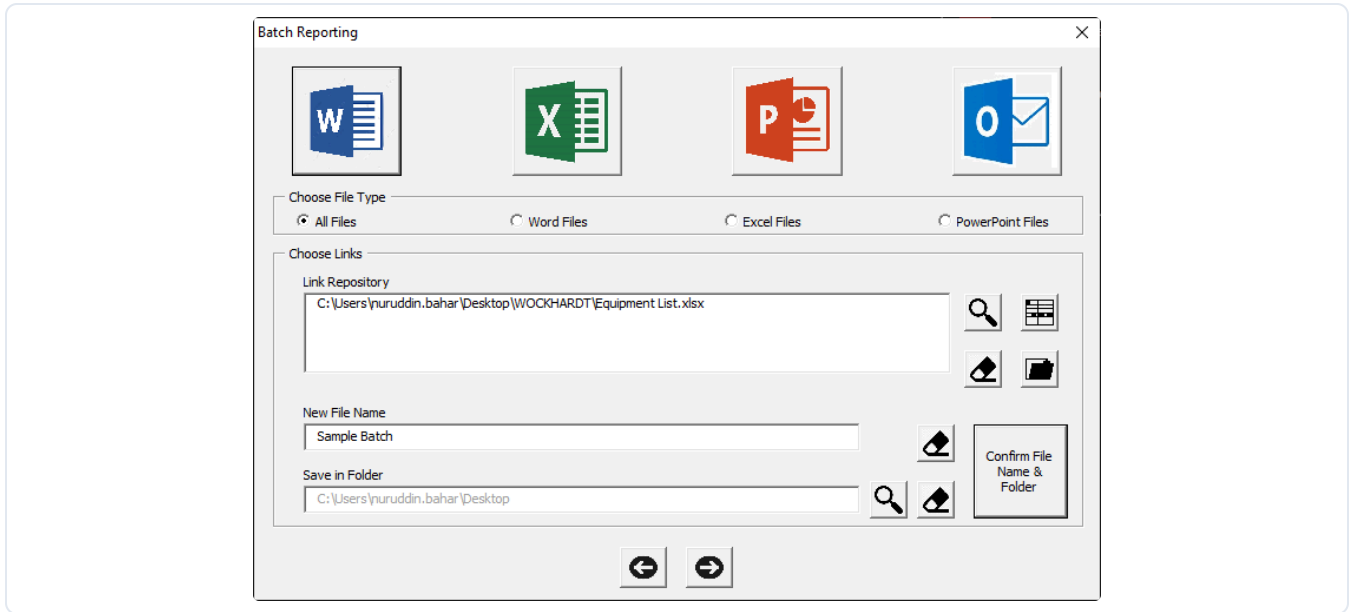


FIG. 03 Reporting wizard — Word / Excel / PowerPoint output

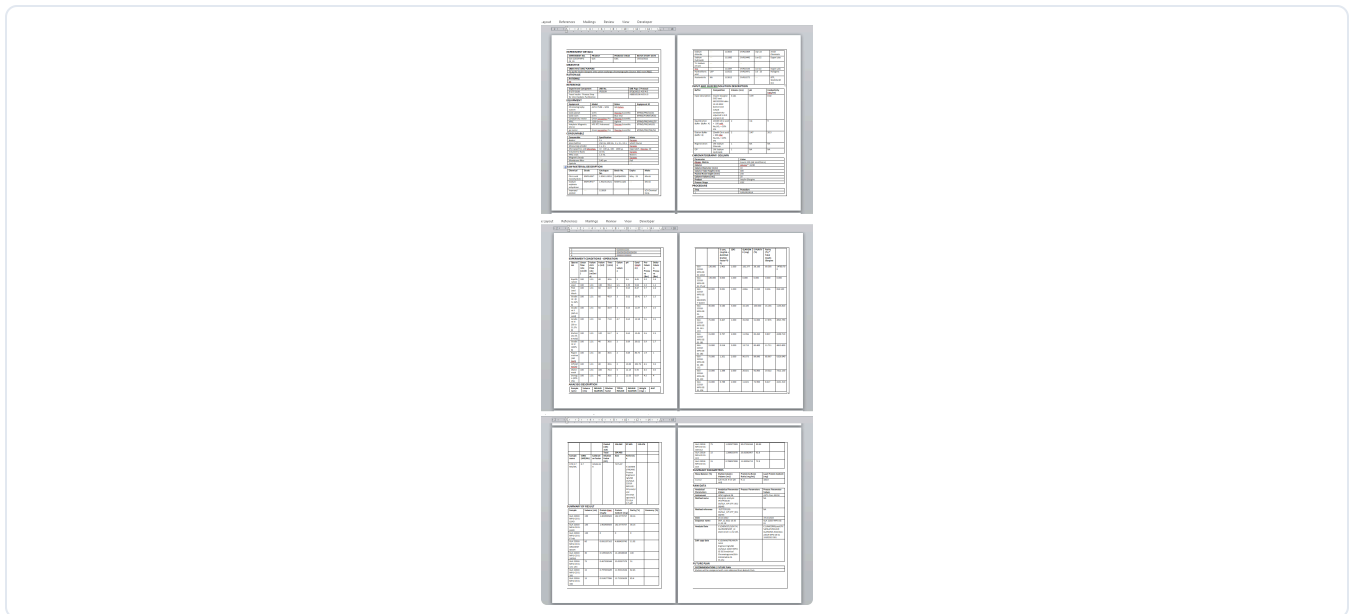


FIG. 04 Generated batch reports compiled automatically across Word, Excel, and PowerPoint

# Statistical Analysis for Process Development

## OVERVIEW

A set of applied statistical studies supporting biologics process development. In the lead study I evaluated alternate-vendor TBEE reagent for the pro-insulin-to-insulin-ester conversion across four experiments.

Using prediction profiling, regression, and equivalence testing to compare standard, alternate, and recovered reagent, I identified a ~98.5% conversion optimum and statistically confirmed batch-to-batch consistency — turning lab data into defensible decisions.

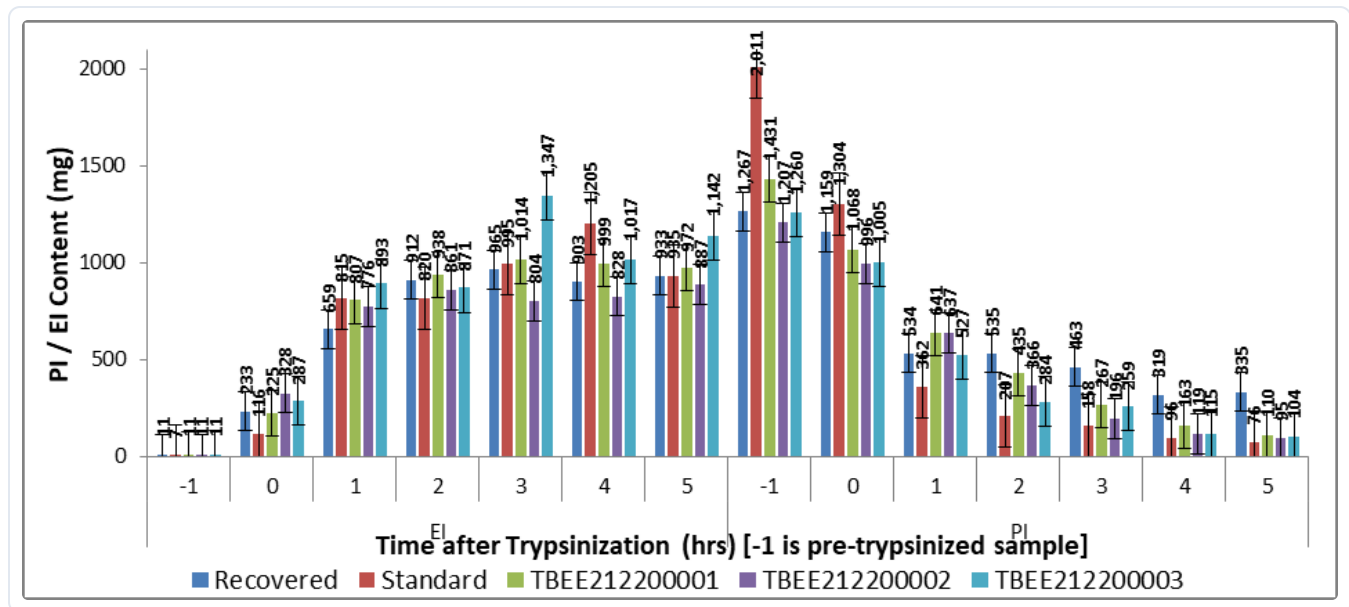


FIG. 01 Pro-insulin to ester-insulin conversion content

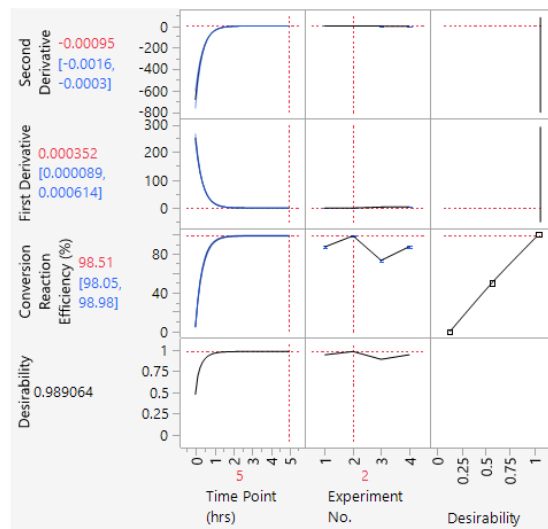


Figure 5.3.1-4: Prediction Profiler for the exponential 3P model for standard TBEE samples

FIG. 02 Prediction profiler optimising conversion efficiency



# Impurity Analysis of Analytical Data

## OVERVIEW

A reusable workflow for analysing chromatographic impurity data across Excel, Python, and Power BI. I clean raw analytical exports, compute relative retention times (RRT) in Python, and present the results as 100%-stacked pivot views that make impurity profiles easy to compare across samples.

It standardises a common QC task and bridges spreadsheet, code, and BI tooling — from raw data to interactive, comparable impurity profiles.

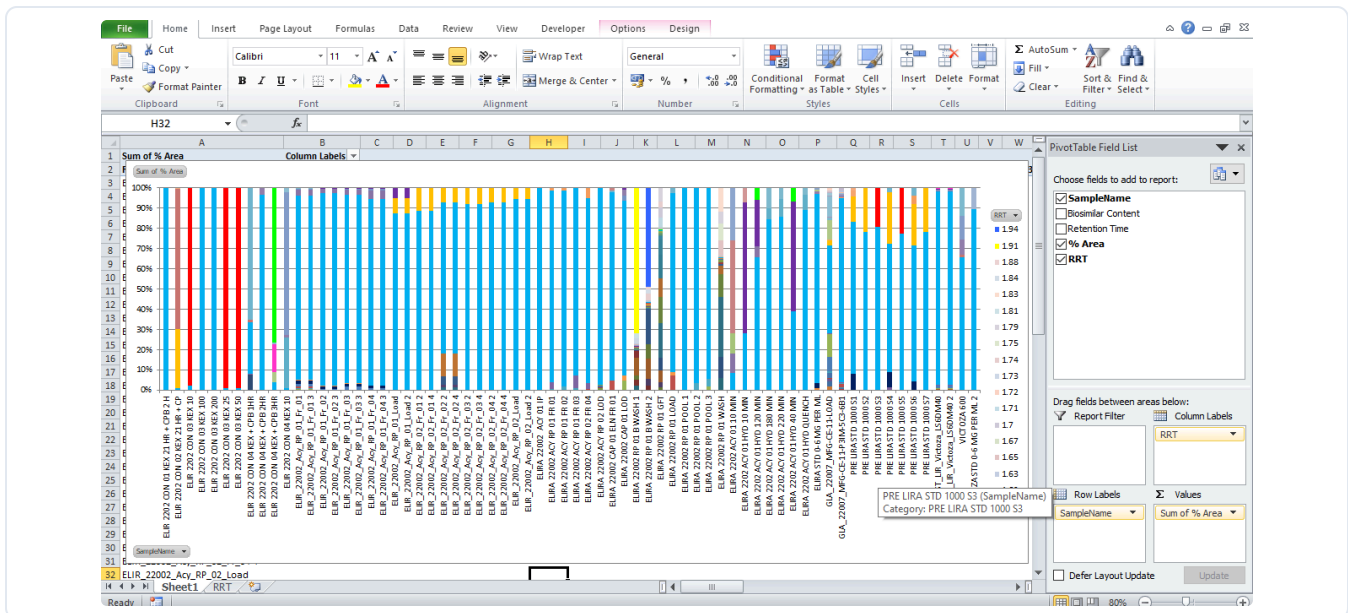


FIG. 01 100% stacked impurity pivot chart

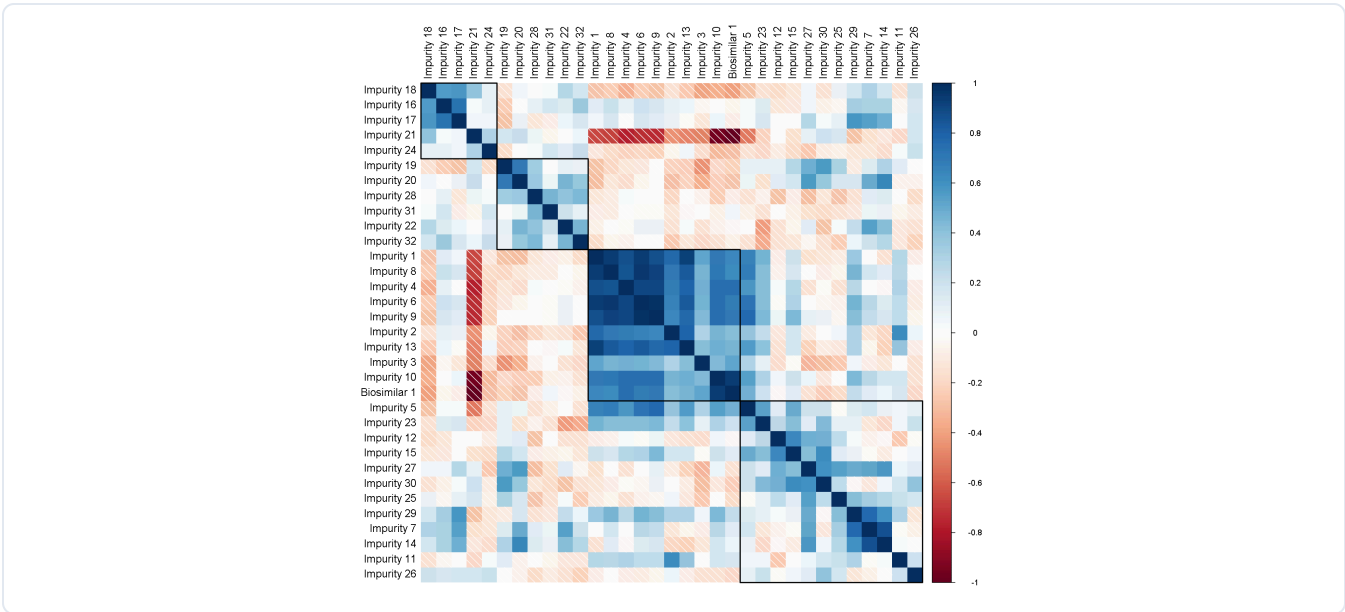


FIG. 02 Impurity correlation heatmap



# Design of Experiments (DOE) GUI

## OVERVIEW

A Python/tkinter application that generates Design-of-Experiments plans from a simple parameter table. The user supplies factors with end-point and/or central values and picks one or more algorithms — full and fractional factorial, Plackett-Burman, Box-Behnken, Latin Hypercube, Sukharev grid, and others.

The tool returns an optimised experiment matrix ready to run, putting rigorous experimental design within reach of non-statisticians and cutting planning time.

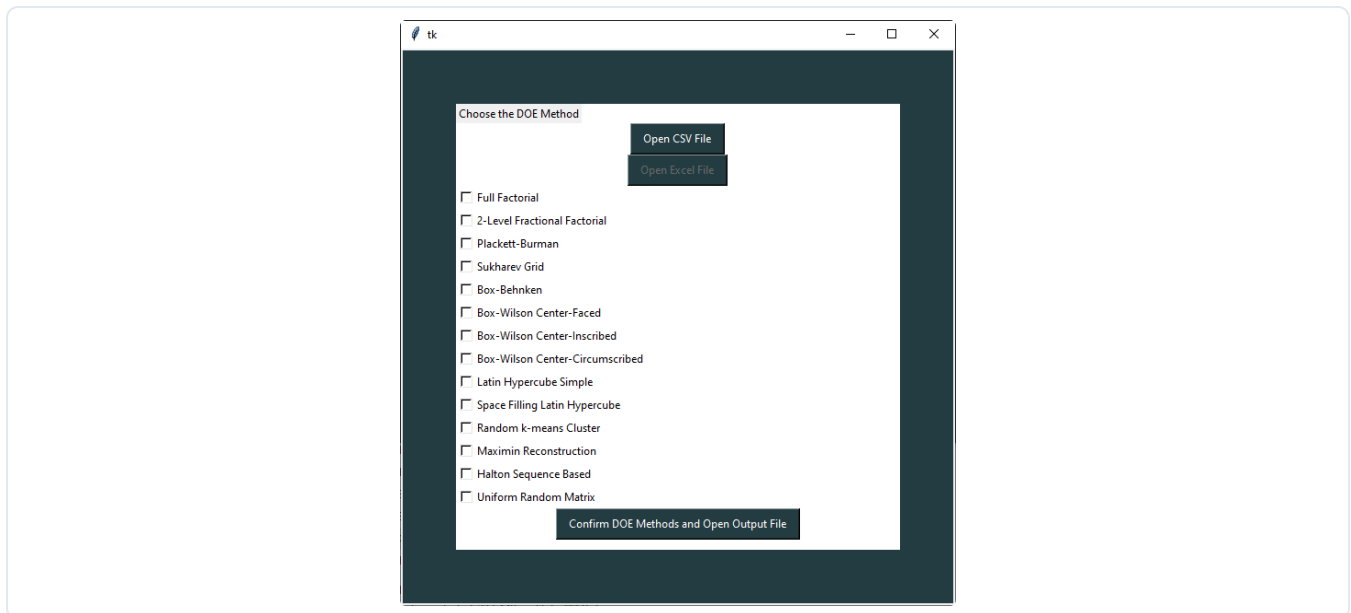


FIG. 01 The DOE GUI — choose one or more design algorithms

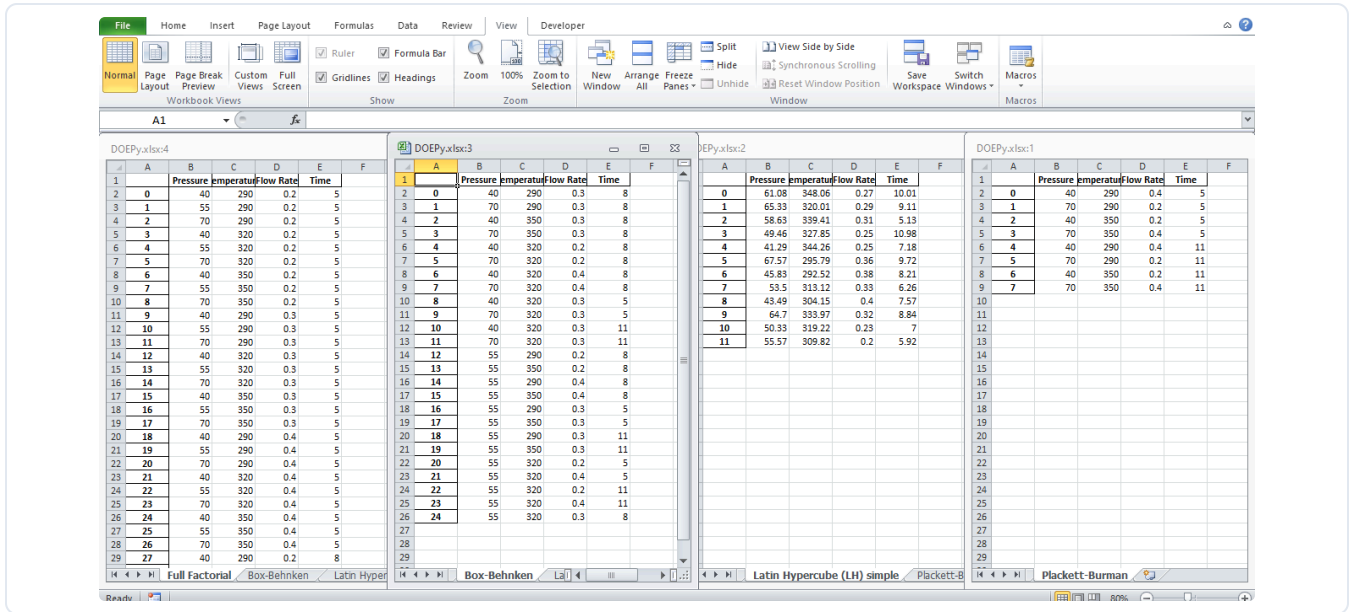


FIG. 02 Generated DOE tables with optimal experiment runs



# Python Resume Maker

## OVERVIEW

A self-contained resume generator that renders a print-ready A4 layout from an Excel content table, using matplotlib's annotation system for precise text placement.

Content, styling, and layout parameters live in separate worksheets, so the design is fully data-driven and reusable — a creative, low-cost solution that doubles as a demonstration of programmatic document layout.

**NURUDDIN BAHAR**  
MS, MTech, BTech

**EDUCATION**

**MS (Biomedical Engineering)** 2018 - 2020  
University of Glasgow, Glasgow, UK  
• Qualified with Distinction  
• THESIS: Design and characterization of poly(ethylene glycol) laminin (PEG LM) based mechano responsive hydrogels for drug delivery and gene therapy Grade A4 (19/22)

**MTech (Nanotechnology)** 2015 - 2017  
Indian Institute of Science, Bangalore, India  
• Qualified with CGPA 5.8/8.0  
• THESIS: Towards DNA Origami Based Magnetic Nanopropellers Grade-A (7/8)

**BTech (Chemical)** 2009 - 2013  
Indian Institute of Technology Bombay, Mumbai, India  
• Qualified with CPI 7.53/10.0

**WORK EXPERIENCE**

**Research Scientist**  
Wookhardt Research Centre, Aurangabad, India  
Bioprocess Engineering and MSAT 10/2020 - Present  
• In process of creating a Django website to display relevant projects  
• Created a VBA-based robotic batch process recording application for ease of reporting  
• Statistically calculated mock pools for CIEC stage of Insulin Glargine and RP stage of Yeast Human Insulin using pandas and matplotlib  
• Created a Python-based design of experiments user interface for creating experimental plans for given parameters  
• Statistically analyzed 2021-22 batches of Yeast Insulin, applied Python-based data visualization, inferential statistics and descriptive statistics  
• Built Gantt charts, timeline charts, chromatograph visualization tool using Python

**Associate Software Engineer**  
Exeter Software Pvt. Ltd., Bengaluru, India 07/2013 - 01/2014  
• Learned essentials of Javascript, SQL, HTML, CSS and Salesforce as a part of training  
• Certified with Oracle Policy Automation 10 Rule Developer Essentials 120-534  
• Created an online passport application portal using Salesforce

**TRAINING AND INTERNSHIPS**

**Skill Development Trainee**  
Centre for Cellular and Molecular Biology, Hyderabad, India 10/2017 - 01/2018

Aurangabad (MAH), INDIA  
rbahar2@gmail.com  
(+91) 701-436-8446  
www.musinginbiology.com

**Scholastic Achievements & Awards**

- AIR 718 - IITEE 2009
- AIR 16 - GATE (Biotechnology) 2015
- CSIR-UGC (Life Sciences) 2015, 2018
- CBSE 10th merit certificate for 100% in social science 2006

**Certifications**

- Certified Associate in Python Programming (Python Institute)
- MS Excel Expert (Microsoft)
- MS Word Expert (Microsoft)
- MS Outlook Associate (Microsoft)
- MS PowerPoint Associate (Microsoft)

FIG. 01 Generated resume rendered with matplotlib

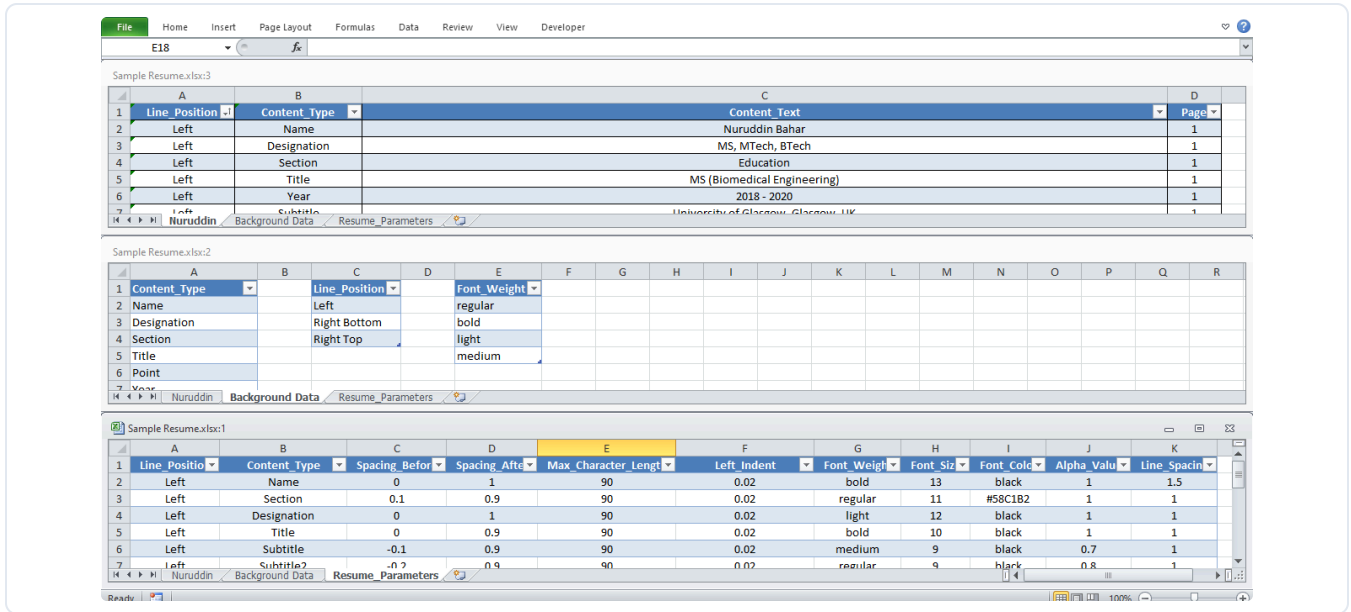


FIG. 02 Excel content and parameter tables driving the layout



# Django Website Platform

## OVERVIEW

A full-stack Django platform where authenticated users publish and share articles on data analysis, visualisation, biology, and Python tooling. It combines custom HTML/CSS/JS front-end work with Django's back end.

Features include article management (create / edit / publish, comments, likes, social sharing) and a browsable project catalogue — demonstrating end-to-end web development from data model to deployed UI.



FIG. 01 Home page with article feed and project sidebar

Home About Logout Projects Search

**Recent Projects**

- Biology
- CNA Transcription and Translation
- Python Tools
- Resume Maker

**Projects by Week**

- May 15 - May 21 (5)
- Apr 17 - Apr 23 (5)

**Chromatogram Visualization Tool**

The chromatogram visualization tool built with Python will help users in viewing chromatographic data in an interactive way using just an Excel / CSV file.

by: [Nuruiddin Bahar](#) updated: May 27, 2023

**Motivation**

As a Process Scientist in biotechnology, I have learned how to generate batch reports for process chromatogram data. Chromatography systems can be simple or complex, with some having an interface to monitor relevant parameters during a process run, while others generate a CSV/Excel file at the end of the run. The report typically displays two parameters on the chromatography plot, which are placed on each vertical axis of a timeline graph.

To get a complete picture of the run, process scientists often have to create multiple graphs and correlate time points, CV points or volume[mL] points. This can be time-consuming and counter-productive, as some scientists even have to manually insert required details in a PowerPoint image.

I have found that Python has useful tools that can help generate a plot with all necessary parameters on a single graph without affecting the results of another. In this article, I will show you

FIG. 02 Article detail page with comments and sharing



# Bioprocess Risk Assessment Tool

Guided Excel / VBA FMEA

---

## ABSTRACT

A guided Excel/VBA application that runs a structured FMEA over a biopharmaceutical downstream process. A standardized questionnaire maps each parameter's answers to fixed Severity, Occurrence, and Detectability ratings via a master matrix, multiplies them into a Risk Priority Number, and ranks unit operations with a Pareto cut-off — with an Ishikawa fishbone generator for root-cause review. Around 5,000 lines of VBA (dynamic forms, keystroke-level input validation, password protection, user management) turn a slow, subjective review into a fast, auditable, ICH-Q9-aligned instrument.

## STANDARDIZED RATING MATRIX

Severity, Occurrence, and Detectability each have a fixed rating scale tied to plain-language criteria, with Detectability inverted so that 'not detectable' carries the highest rating. A master lookup converts each answered question into its rating, so scoring is consistent and traceable to the criteria rather than to an assessor's gut feel.

## RPN & PARETO PRIORITISATION

Each parameter's RPN is Severity × Occurrence × Detectability, and a pivot sums RPN by unit operation and sorts it descending with a running cumulative percentage. An 80% cut-off marks the 'vital few' — in this run the top four risk items, led by Virus Inactivation and the chromatography and filtration steps, account for roughly 80% of total risk — focusing mitigation where it actually moves the needle.

## FORMS, VALIDATION & AUTOMATION

The workbook is a full VBA application — about 5,000 lines across a dozen UserForms and a set of helper classes. Custom class modules enforce numeric, alphanumeric, and synchronized-dropdown input at the keystroke; a dynamic form builder generates the entry screens from the workbook's own tables; and further modules handle submission, sheet visibility, charting, password protection, and user management.

## IMPACT

It turns a slow, subjective review into a fast, standardized one that any trained assessor runs the same way — results ranked, auditable, and ready for root-cause follow-up. Its questionnaire-to-RPN mapping bakes ICH Q9 risk thinking into the tool, keeping decisions consistent across parameters and assessors.

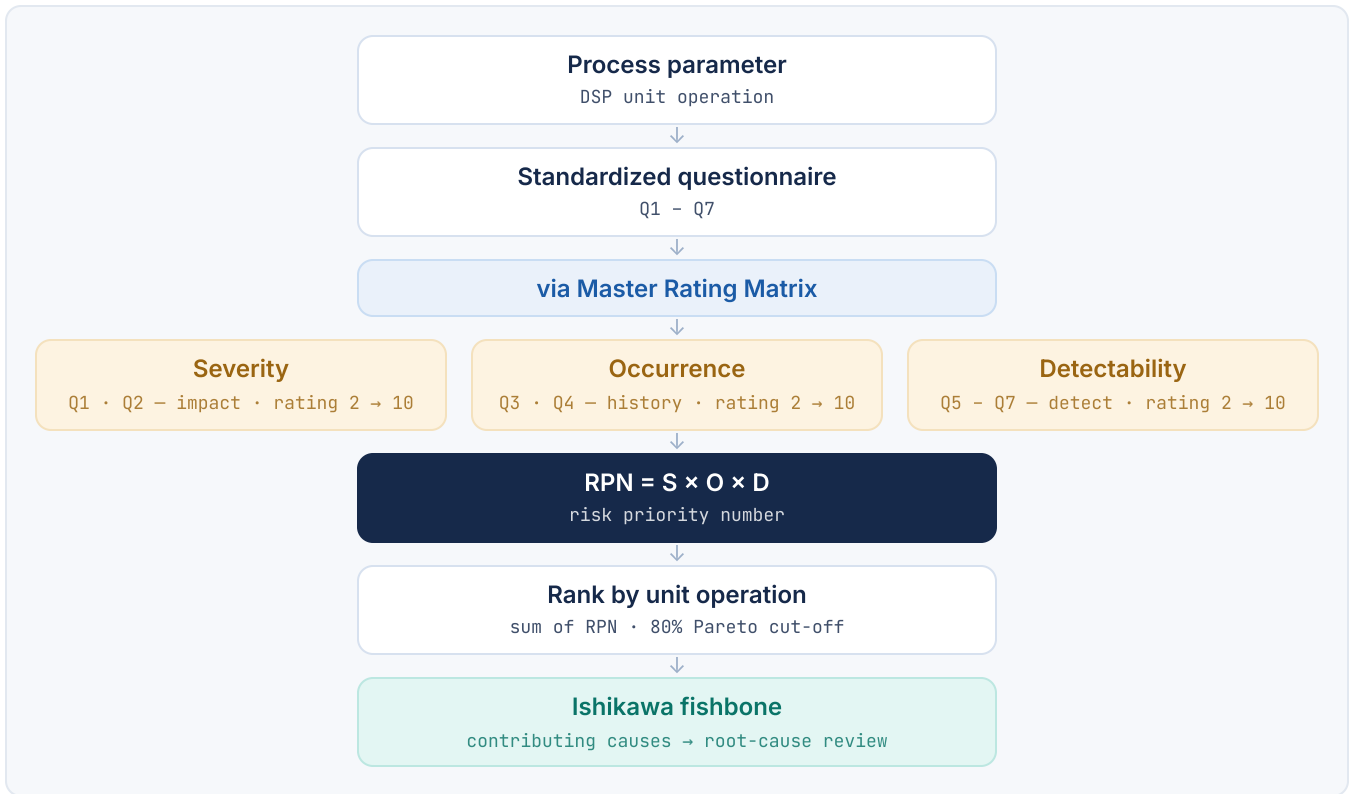


FIG. 01 Scoring model — a questionnaire maps to fixed S, O & D ratings, multiplied into RPN, then Pareto-ranked and traced to causes

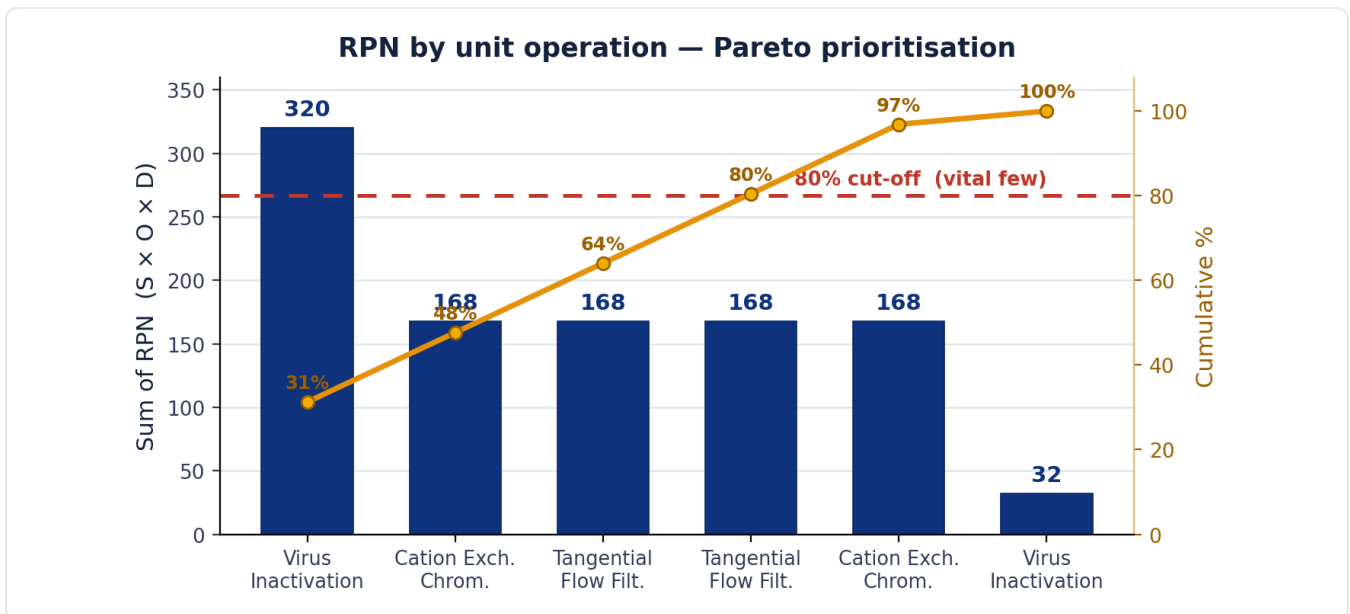


FIG. 02 The fixed S / O / D rating scales (top) and the resulting RPN Pareto that prioritises the highest-risk unit operations (bottom)



# Golden Batch Analysis

## Multivariate Bioprocess Analytics

---

### ABSTRACT

A multivariate analytics pipeline that mines a biologic's manufacturing history to find the process profile behind its best batches. Working from multi-campaign Abatacept 1KL data, it imputes missing values (KNN), reduces dozens of correlated parameters with PCA, models a quality attribute with Partial Least Squares regression, then isolates 'golden' batches by threshold and flags anomalies with Mahalanobis and Cook's distance. Built as a modular pipeline with reusable function libraries and a small GUI, it supports process characterisation and continued process verification by turning routine batch data into guidance for future production.

### DIMENSIONALITY REDUCTION & EXPLORATION

With dozens of correlated upstream parameters, the pipeline standardizes everything and runs PCA, using a Pareto chart of cumulative explained variance to keep only the components that carry signal. A biplot of the first two components, normalized box plots, and a per-component maximum-correlation step make the structure interpretable — showing which raw parameters drive each component before any model is fit.

### PREDICTIVE MODELLING (PLS)

A Partial Least Squares regression predicts the quality attribute from the selected parameters, trained on an 80/20 split and reported with  $R^2$  and RMSE. PLS is chosen over ordinary regression precisely because the predictors are collinear, with VIF confirming the multicollinearity it is built to handle. On the golden-batch subset the model tracks the attribute closely, giving a parameter-based predictor of batch quality.

### GOLDEN BATCHES & OUTLIERS

Batches above a target-quality threshold are tagged as golden, exported for reference, and compared with 'golden-like' batches through the mean absolute difference of their principal-component errors — quantifying how far each batch drifts from the ideal. Separately, Mahalanobis and Cook's distance flag multivariate outliers, computed with and without time-series treatment so unusual runs surface whichever way the data is viewed.

### IMPACT

It turns a large, messy manufacturing record into a defensible answer to a practical question — what does a great batch look like, and which deviate? The result is a reusable pipeline for process characterisation, turning routine batch data into guidance for steering future production toward the golden profile.



FIG. 01 Analysis pipeline — impute, reduce with PCA, model the quality attribute with PLS, then isolate golden batches and outliers

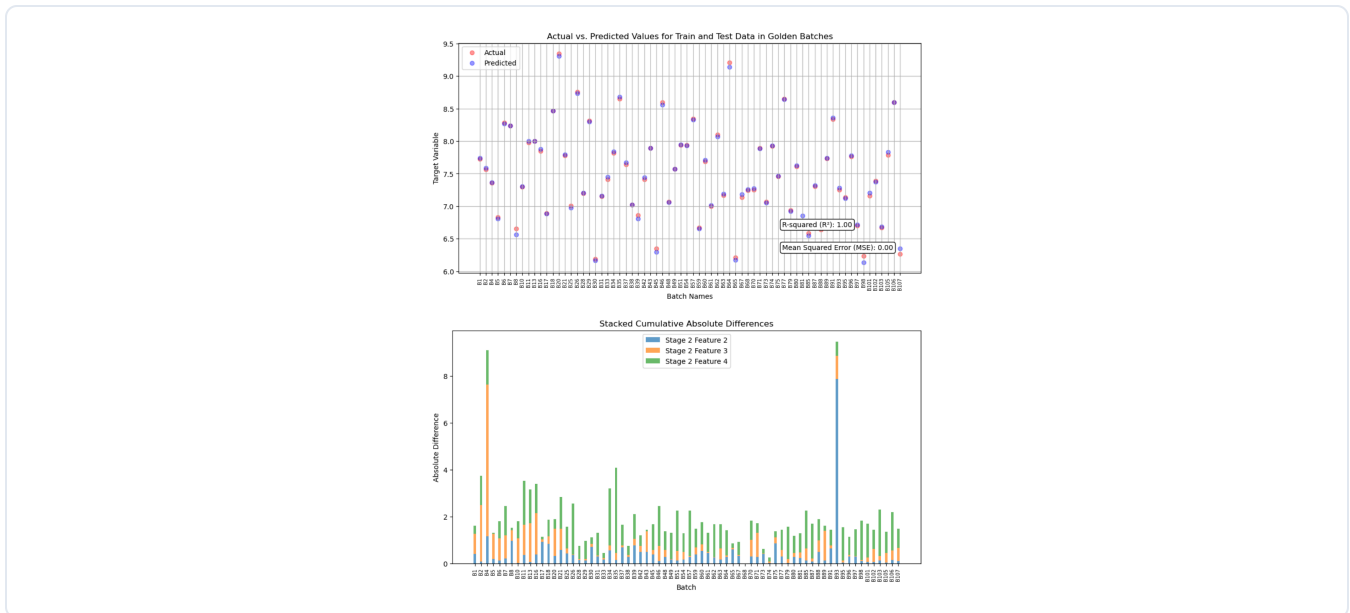


FIG. 02 PLS actual-vs-predicted quality for golden batches (top) and per-batch deviation from the golden profile by principal component (bottom)

# DNA Origami Magnetic Nanopropellers

## Structural DNA Nanotechnology

### OVERVIEW

My M.Tech thesis developed DNA-origami-based magnetic nanopropellers — self-assembling nanoscale structures designed to move through biological media under a magnetic field. I optimised iron-oxide ( $\text{Fe}_3\text{O}_4$ ) nanoparticle synthesis by co-precipitation, designed six-helix bundle origami in caDNAno, characterised the assemblies by atomic force microscopy, and bioconjugated the magnetic particles to the DNA scaffold.

The work targets magnetically steerable nanocarriers for drug delivery and produced a conference paper — demonstrating hands-on nanotechnology, structural DNA design, and rigorous experimental characterisation.

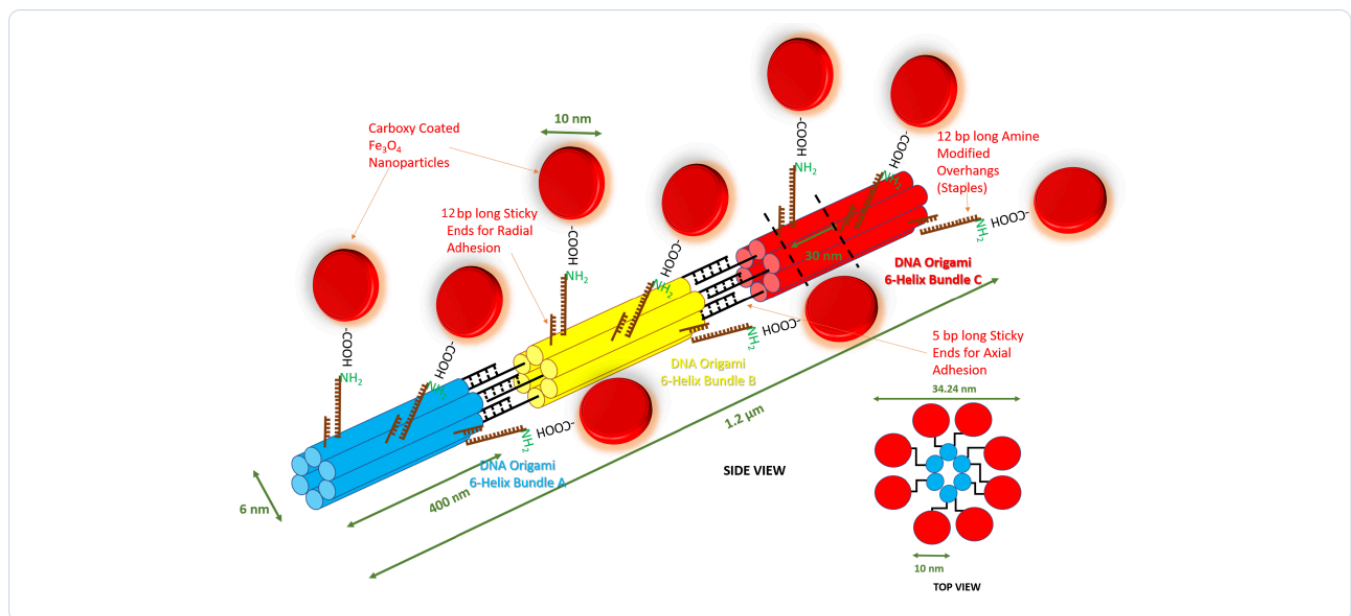


FIG. 01 Geometric reconstruction of the DNA origami nanopropeller

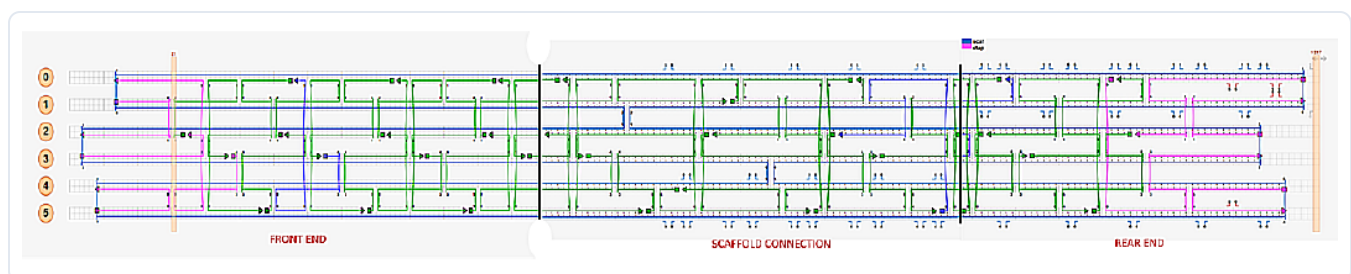


FIG. 02 caDNAno design of the six-helix bundle

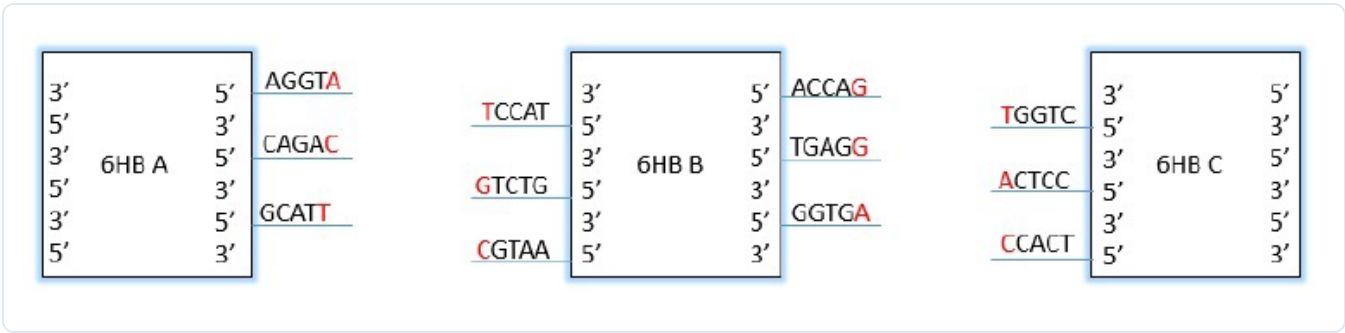


FIG. 03 Sticky-ended cohesion forming bundles A + B + C

# Mechanoresponsive PEG-Laminin Hydrogels

## OVERVIEW

For my MSc I engineered mechanoresponsive PEG-laminin hydrogels seeded with human mesenchymal stem cells, to study how a material's mechanics steer cell behaviour for tissue regeneration. By tuning the ratio of 4-arm PEG cross-linker to an MMP-degradable VPM peptide, I controlled stiffness for vascular and bone applications.

Using bulk rheology and a novel nanoindentation method, I showed the bulk shear modulus depends on applied normal force rather than chemical composition — pairing biomaterials design with quantitative mechanical characterisation.

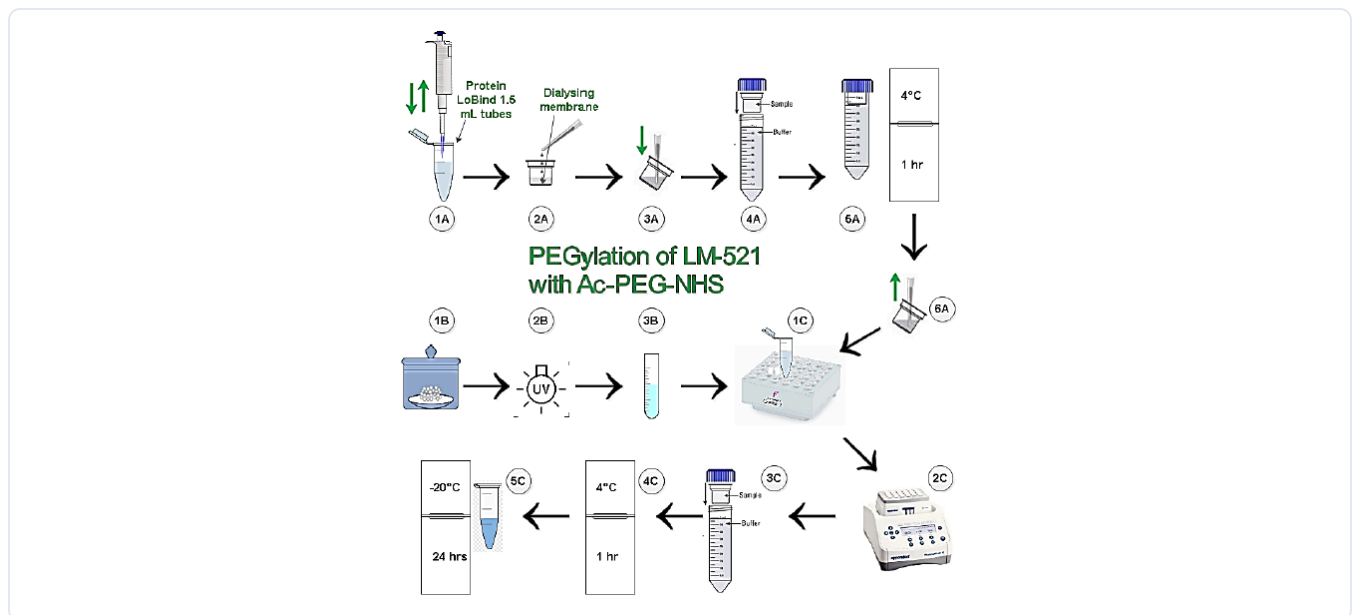


FIG. 01 PEGylation of LM-521 with Ac-PEG-NHS

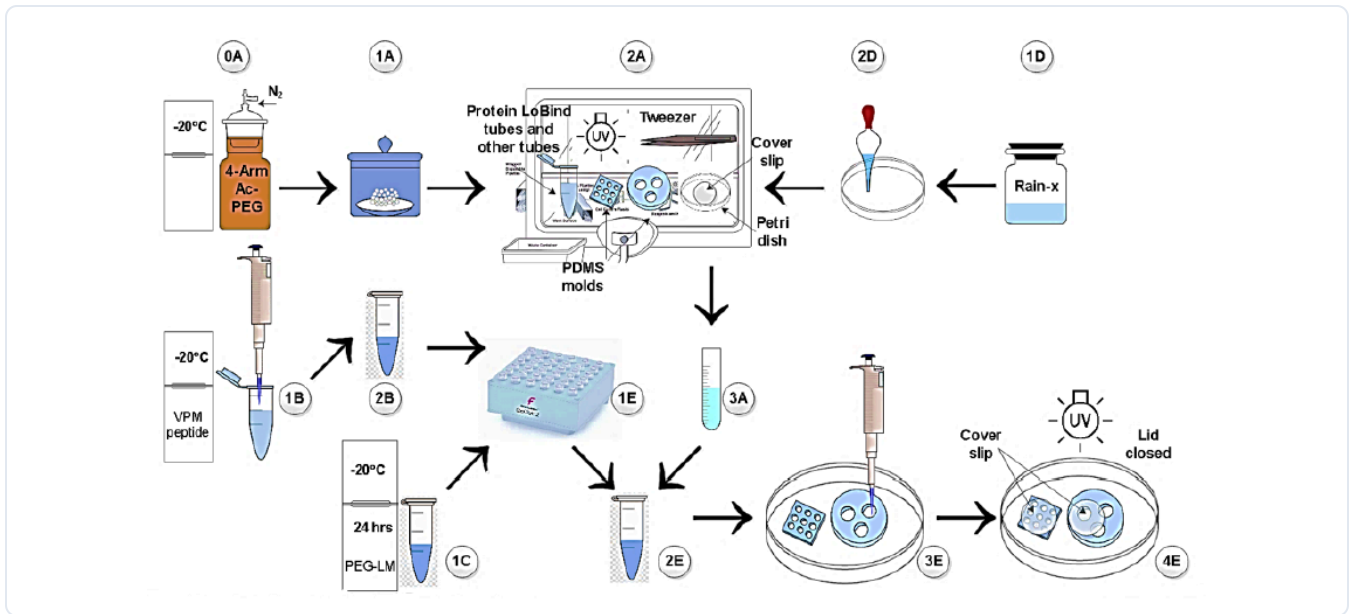


FIG. 02 Hydrogel fabrication and characterisation workflow



# Scientific Review Papers

## OVERVIEW

Authored scientific review papers that translate dense primary literature into structured, accessible syntheses — including a review of ocular tissue engineering covering stem-cell and scaffold-based approaches to vision restoration, their clinical status, and open challenges.

The work demonstrates literature mastery, critical synthesis, and clear technical writing across biomedical topics.

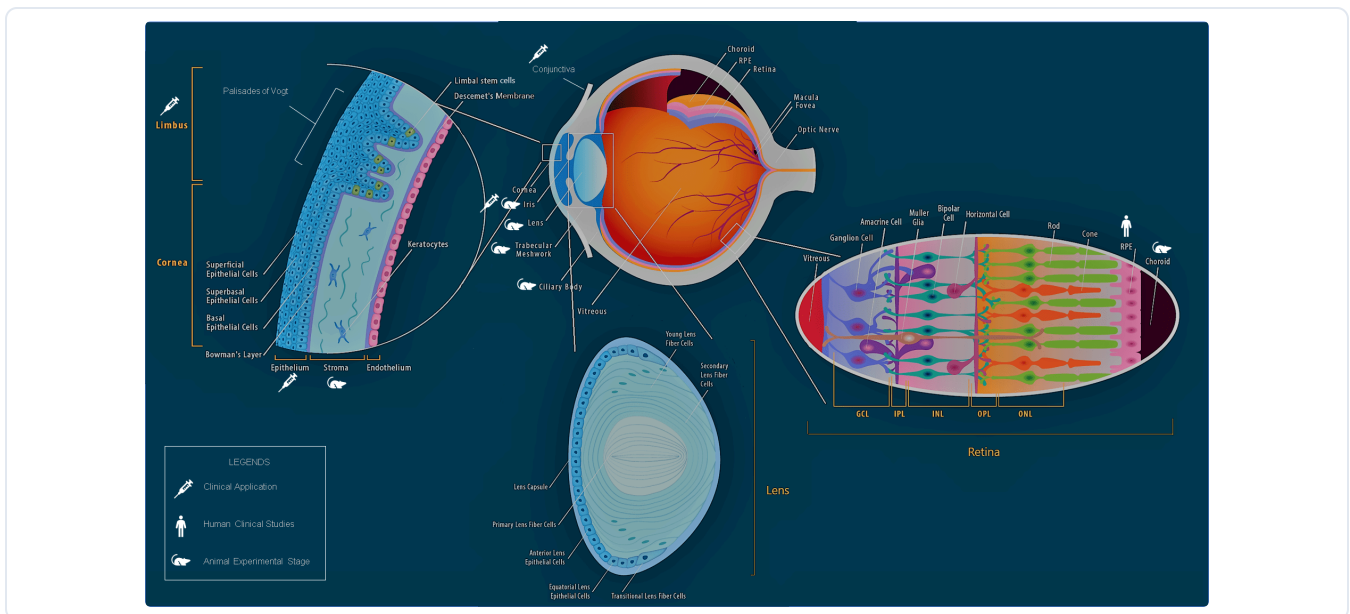


FIG. 01 Eye anatomy — from the ocular tissue-engineering review

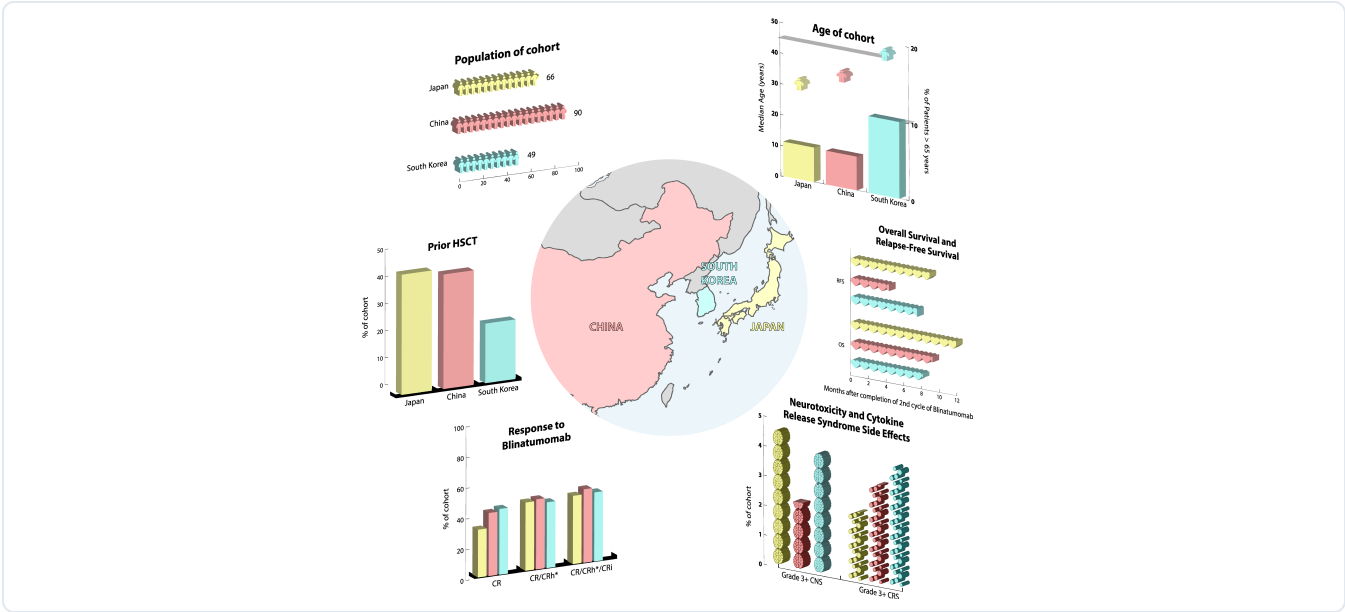


FIG. 02 Figure from a second review paper