

# Leveraging Integrated Data Warehousing and Interactive Dashboards to Monitor Success Patterns in Assisted Reproduction

Abhimanyu C. J.<sup>1</sup>; Kiran Kapuram<sup>2</sup>; Challa Venkatshevarlu<sup>3</sup>; Gayathri K.<sup>4</sup>; Bhargavi Depuru<sup>5</sup>; Srija Depuru<sup>6</sup>; Bharani Kumar Depuru<sup>7</sup>

<sup>1,2,3</sup>Research Associate, <sup>4</sup>Project Mentor, <sup>5</sup>Team Lead, <sup>6,7</sup>Director

<sup>1,2,3,4,5,6,7</sup> AISPRY Technologies Pvt Ltd.

Publication Date: 2026/03/09

**Abstract:** Precise tracking of In-Vitro Fertilization (IVF) procedures is vital for improving medical results and refining patient guidance. Nevertheless, the intricate nature of multidimensional clinical datasets, which include diverse patient profiles and medical protocols, frequently prevents fertility centers from obtaining a comprehensive understanding of success trends. This study introduces "Fertility Insights," an analytical approach utilizing the CRISP-ML(Q) methodology to simplify IVF oversight through data-centered visualization. The architecture employs a local data stream where primary clinical files are refined via Python for stringent data validation and normalization. The cleaned data is then hosted in a unified MySQL repository, maintaining data consistency and organized retrieval for longitudinal studies. Lastly, dynamic dashboards are created within Power BI to illustrate essential performance metrics, including the variance in success rates. This system equips healthcare professionals with practical intelligence regarding therapeutic impact and facility productivity. This research highlights how descriptive modeling can effectively turn complicated reproductive data into a functional resource for clinical strategy.

**Keywords:** IVF Monitoring, Descriptive Analytics, Data Visualization, Power BI, MySQL Data Warehouse, Fertility Insights, Clinical Data Management, Patient Demographics, Treatment Success Analysis, Growth Hacking in Healthcare.

**How to Cite:** Abhimanyu C. J.; Kiran Kapuram; Challa Venkatshevarlu; Gayathri K.; Bhargavi Depuru; Srija Depuru; Bharani Kumar Depuru (2026) Leveraging Integrated Data Warehousing and Interactive Dashboards to Monitor Success Patterns in Assisted Reproduction. *International Journal of Innovative Science and Research Technology*, 11(2), 2953-2963. <https://doi.org/10.38124/ijisrt/26feb1461>

## I. INTRODUCTION

Infertility has emerged as a significant worldwide health concern, impacting nearly 15% of reproductive-aged couples [1]. In the realm of contemporary Assisted Reproductive Technology (ART), the industry is transitioning toward a "P4 medicine" model—an approach defined by preventative, individualized, inclusive, and foresight-driven care [1]. This strategy utilizes expansive datasets and sophisticated machine learning to combine lifestyle, genetic, and ecological variables to navigate the complex biological hurdles of pregnancy. Even with these advancements, In-Vitro Fertilization (IVF) success probabilities have plateaued at roughly 35% per embryo transfer, highlighting the need for more dependable, data-centric oversight. A central clinical and operational challenge within IVF facilities is the heavy dependency on qualitative embryo morphology reviews. Conventional manual grading is naturally prone to human oversight, leading to a notable 25–40% disparity in embryo classification among different specialists. Such variability

results in less-than-ideal embryo prioritization, diminished pregnancy outcomes, and a marked rise in logistical and financial burdens for both providers and patients [2]. At present, many fertility clinics lack a uniform, analytical strategy to pinpoint the embryos with the highest conception potential, frequently forcing medical staff to depend on intuitive risk assessments and probability [1].

The goal of this investigation is to conduct an initial ad-hoc evaluation focused on increasing ART success metrics by 10% through descriptive analytics applications. By improving how data is interpreted, this research intends to lower patient expenses while enhancing the functional productivity of clinical teams [2]. To create a "Unified Data Authority," we suggest a technical architecture featuring a data pipeline that employs Python for cleansing and organization, MySQL for central storage, and Power BI for live tracking and graphical reporting. This structure is designed to connect patient-sourced health metrics with

clinical support tools to facilitate more reliable and favorable fertility results [1,2].

## II. METHODOLOGY AND TECHNIQUES

This investigation employs the Cross-Industry Standard Process for Data Mining (CRISP-DM) as the foundational methodology for the development and execution of the Fertility Insights system. CRISP-DM is an organized, recursive framework recognized globally for steering data science initiatives from the initial concept through to final implementation [3]. Its six interconnected stages—Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment—offer a stable blueprint for handling intricate medical datasets while ensuring rigorous oversight and process documentation [4].

### ➤ *Business Understanding*

In the IVF sector, the main goal is to optimize procedural tracking and increase clinical success through organized descriptive modeling. During this stage, vital performance metrics such as embryo transfer success, fertilization frequency, implantation outcomes, maternal age demographics, and treatment protocol effectiveness were established in cooperation with medical specialists. The clinical difficulty regarding irregular embryo assessment and outcome fluctuations emphasized the requirement for uniform data analysis protocols [5]. The CRISP-DM approach guarantees that research targets stay synchronized with both healthcare and logistical goals, effectively linking clinical expertise with data-driven strategy [6].

### ➤ *Data Understanding*

The dataset consisted of de-identified medical records encompassing patient demographics, endocrine profiles, stimulation strategies, embryological data, and clinical results. In IVF studies, data heterogeneity is a frequent challenge stemming from diverse laboratory procedures and recording formats [7]. Consequently, exploratory data analysis (EDA) was executed through Python to evaluate data integrity, identify outliers, and analyze the distribution patterns of critical variables. Statistical overviews and correlation studies were conducted to interpret the associations between variables and pregnancy outcomes. This stage provided essential transparency and pinpointed structural flaws within the primary clinical information.

### ➤ *Data Preparation*

The data conditioning stage is frequently the most demanding component of medical informatics [3]. Initial IVF data underwent intensive preprocessing, involving the management of missing entries, outlier identification, feature encoding, and scaling where necessary. Redundant entries were eliminated to maintain data reliability. Medical terminology was harmonized to minimize linguistic differences between records. Refined datasets were organized into relational schemas and hosted in a unified MySQL repository to guarantee longitudinal tracking and uniformity. This systematic warehousing method supports continuous outcome tracking and enables expandable analytics within the reproductive health sector [8].

### ➤ *Modeling*

Data modelling While the CRISP-DM framework conventionally incorporates predictive techniques, this investigation focuses on descriptive analytics to fit the research objectives. Data aggregation and statistical assessments were utilized to extract success frequency distributions, age-group classifications, and protocol-based efficiency metrics. These analytical results were not intended for forecasting but served to summarize historical operational trends. This descriptive modeling strategy aligns with modern IVF analytical platforms that emphasize clarity and openness to assist in clinical strategy [9].

### ➤ *Evaluation*

The assessment stage measured the analytical results against the established clinical KPIs. Validation was conducted through cross-referencing with past clinical records and specialist appraisal to confirm reliability and precision. Dashboard prototypes were evaluated for clarity and functional use among reproductive health professionals. In medical data science, verification through the participation of subject matter experts is essential to ensure that data-driven results result in practical clinical intelligence [6,10].

### ➤ *Deployment*

The concluding stage featured implementation via dynamic dashboards created in Power BI. Graphical interfaces were built to display live KPI monitoring, demographic trends, and comparative treatment results. By applying the CRISP-DM cycle within an IVF context, this research illustrates how organized descriptive modeling can transform disjointed medical files into a cohesive oversight platform. This strategy encourages evidence-based clinical management while remaining aligned with medical information security protocols [4,8].

Through the systematic application of CRISP-DM, the Fertility Insights system creates a consistent, high-standard workflow for IVF tracking, balancing technical precision with practical medical utility.

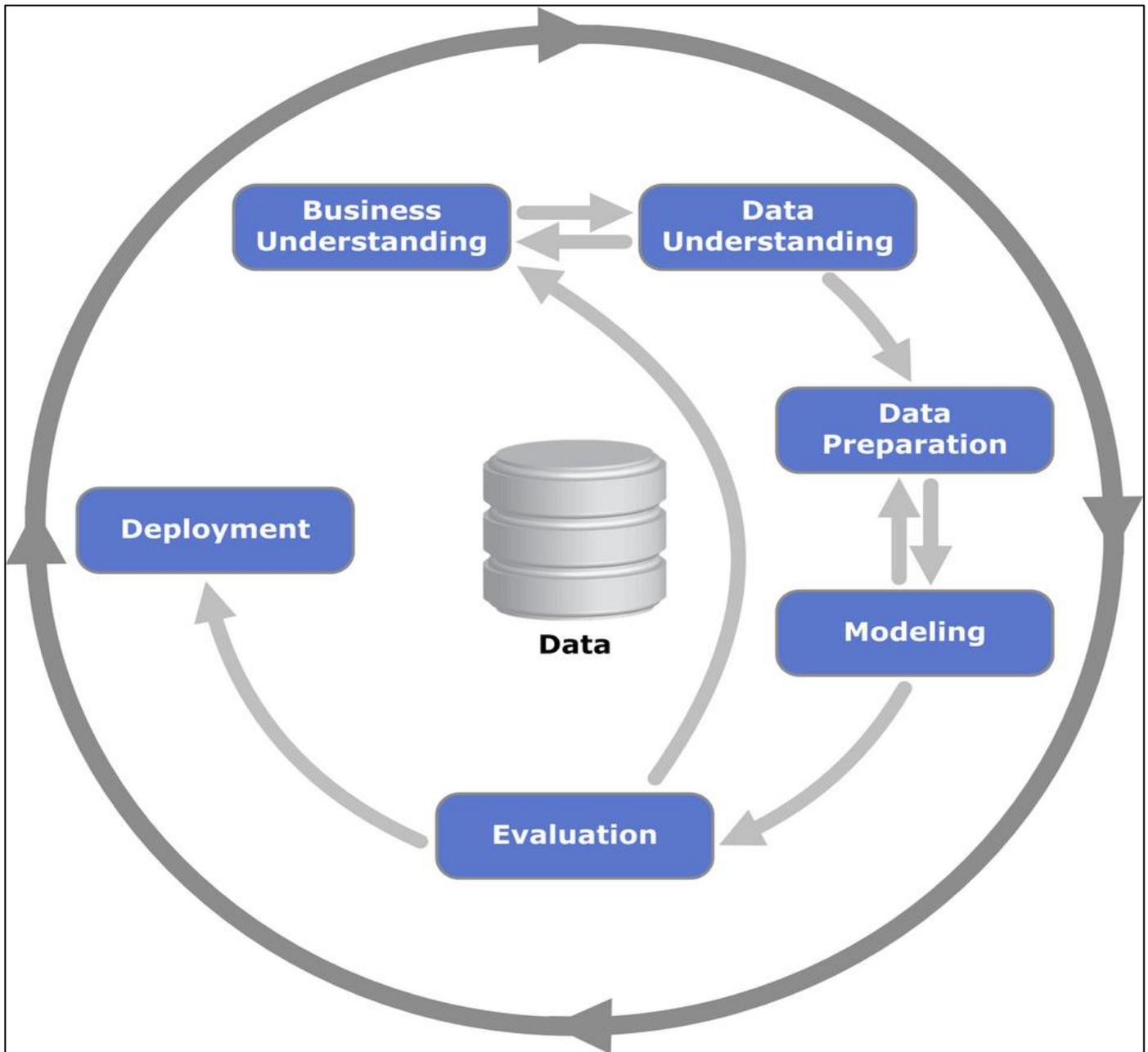


Fig 1 CRISP- DM Methodological Framework, Outlining its Key Components and Steps Visually.  
(Source: -Wikimedia Commons)

### III. ARCHITECTURE

The architecture of this project is structured as a sequential data-driven descriptive analytics pipeline, transitioning from raw data collection to final visualization.

➤ *Data Source (Historical Data):*

The pipeline originates with the collection of Historical Data. This information is stored and accessed through local file formats, specifically Excel and CSV files.

➤ *Data Cleaning (Python):*

The raw data from the source is ingested into Python. This stage is dedicated to Data Cleaning, ensuring that the clinical datasets are refined and standardized for analysis.

➤ *Data Warehouse (MySQL):*

Once cleaned, the data is moved into a MySQL environment. In this architecture, MySQL serves as the centralized Data Warehouse, providing a structured repository for the processed fertility records.

➤ *Data Visualization (Power BI):*

The final stage involves Exporting Data from the MySQL warehouse into Power BI. This component is responsible for the Data Visualization aspect of the project, turning the structured data into monitoring insights.

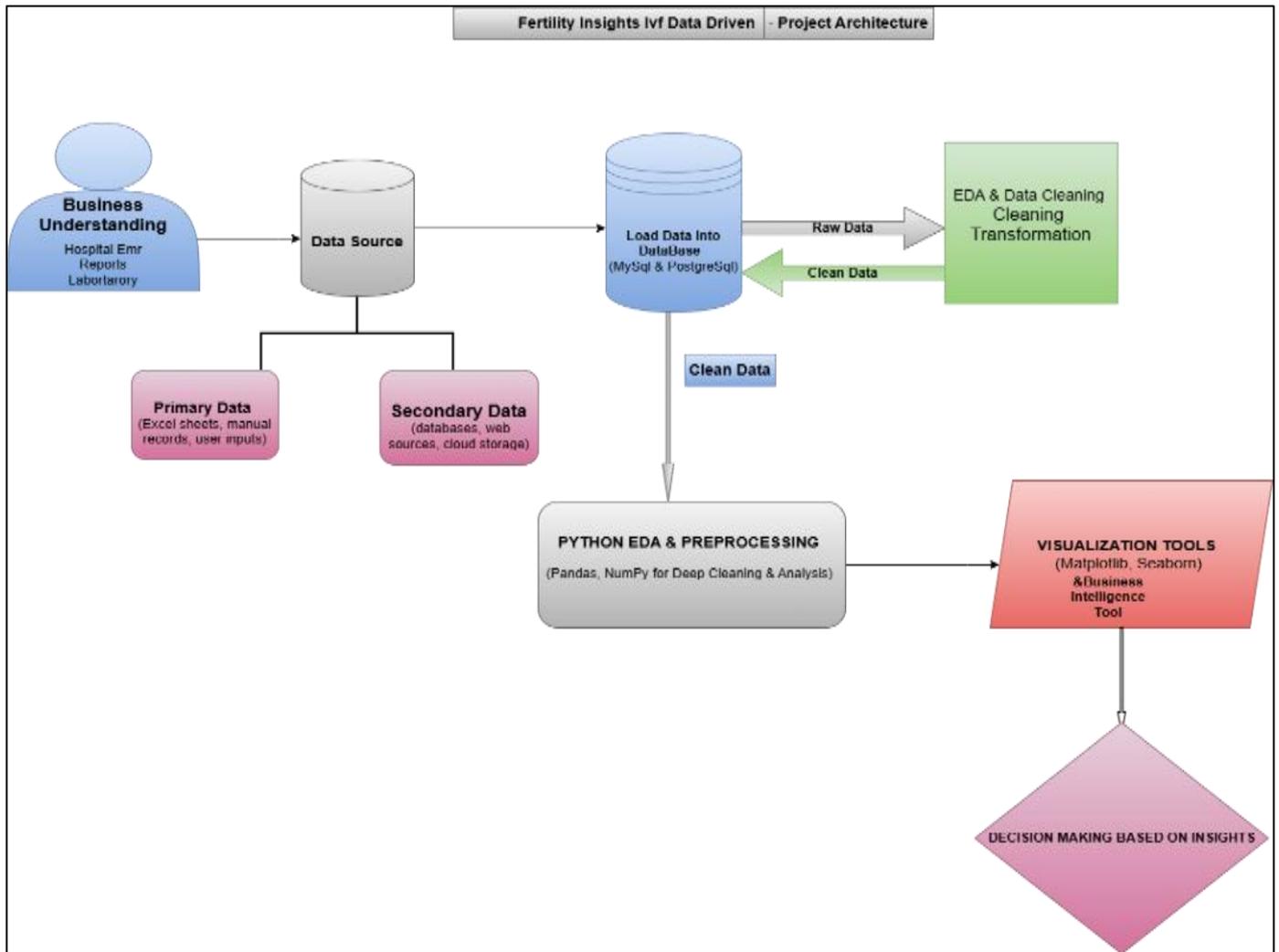


Fig 2 Project Architecture

➤ *Business Understanding*

• *Business Problem and Objectives*

There is significant inconsistency in embryo grading across embryologists (often 25- 40%), leading to suboptimal embryo selection, reduced IVF success rates, longer treatment cycles, and increased operational costs. Clinics lack a standardised, data driven method to identify the embryos most likely to result in successful pregnancies.

Preliminary ad-hoc analysis study to maximise success rate of ART procedure

✓ *Business Success Criteria*

Do a study to Increase success rate of ART procedure by 10 %

✓ *Economic Success Criteria*

Achieve operational efficiency and provide descriptive analytics

✓ *Constraints*

Minimize treatment cost

➤ *Data Understanding*

The study utilizes a structured IVF operational dataset integrating multiple clinical, laboratory, and administrative data sources. The dataset consists of the following relational tables ivf\_appointments, ivf\_equipment, ivf\_operational\_costs, ivf\_patient\_times, ivf\_perforance\_kpis, ivf\_room\_utilization, ivf\_sta. The dataset represents multi-dimensional IVF clinic operations, covering patient scheduling, embryology lab activities, equipment monitoring, cost tracking, staff allocation, and key performance indicators (KPIs).

**IV. EXPLORATORY DATA ANALYSIS (EDA)**

The exploratory analysis revealed operational costs, and resource utilization inefficiencies and other factors are critical determinants influencing ART success rates. The integrated dataset structure supports and operational optimization strategies.

➤ *Descriptive Statistical Analysis*

Descriptive statistics were computed to summarize central tendency and dispersion (as shown in Table 1)

Table 1 EDA

E Exploratory Data Analysis (EDA)					
	Amount	Budget Allocated	Purchase Cost	Kpi Value	Usage Date
Mean	47537.2	54886.12	36411.7051	0.749246	2024.11-02
Median	12977.3	14635.23	26142.46	0.749246	2024-11-28
Mode	6335.5	8273.78	3316.01	0.648	2025-06-09
Variance	49663.5	7148762.23	378023	0.04977	0.24748
Std	22285.9	267381.23	61484.4	0.223108	0.49741
Range	34993.7	4451976	417671.02	2.289	1
Skewness	-26880	-2672297.544	37190.23	-216138	152786
Kurtosis	175.702	172.398	30.4425	12.49641	1.104175

• *Key Observations*

- ✓ High variance in budget\_allocated suggests scheduling inefficiencies.
- ✓ Purchase\_cost exhibit right-skewed distribution.
- ✓ Amount shows clustering around peak IVF cycle periods.

➤ *Central Tendency and Dispersion*

For numerical variables such as Citations and Impact Factor:

- Mean citations – indicates average research influence
- Median citations – more robust against outliers

- Standard deviation – measures citation variability
- Skewness & kurtosis – assess distribution shape

➤ *Distribution Analysis*

Distribution analysis was performed to examine statistical behavior, skewness, dispersion, and normality of key clinical and operational variables across integrated IVF datasets. This analysis supports:

- Selection of appropriate statistical tests
- Development of predictive models
- Detection of outliers and operational drift
- Evaluation of cost variability

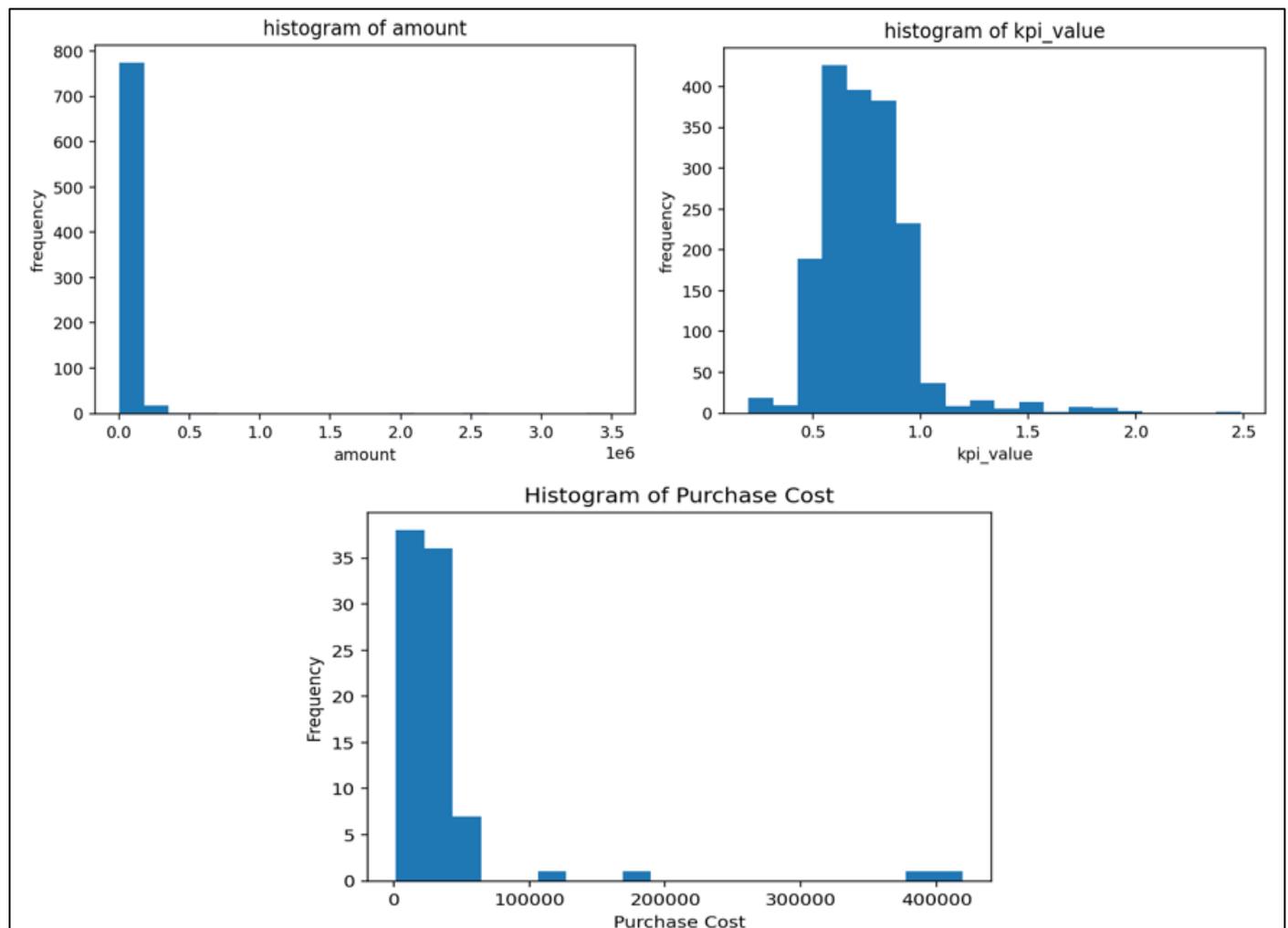


Fig 3 Histogram Analysis

➤ *Histogram Analysis*

Key observations from histogram

- Amount → Right-skewed distribution indicating bottlenecks.

- Purchase Cost → Positively skewed, indicating few high-cost cycles.
- Kpi\_value → The histogram of KPI values reveals a primarily unimodal distribution characterised by a minor positive skew (refer Fig 3- Histogram Analysis)

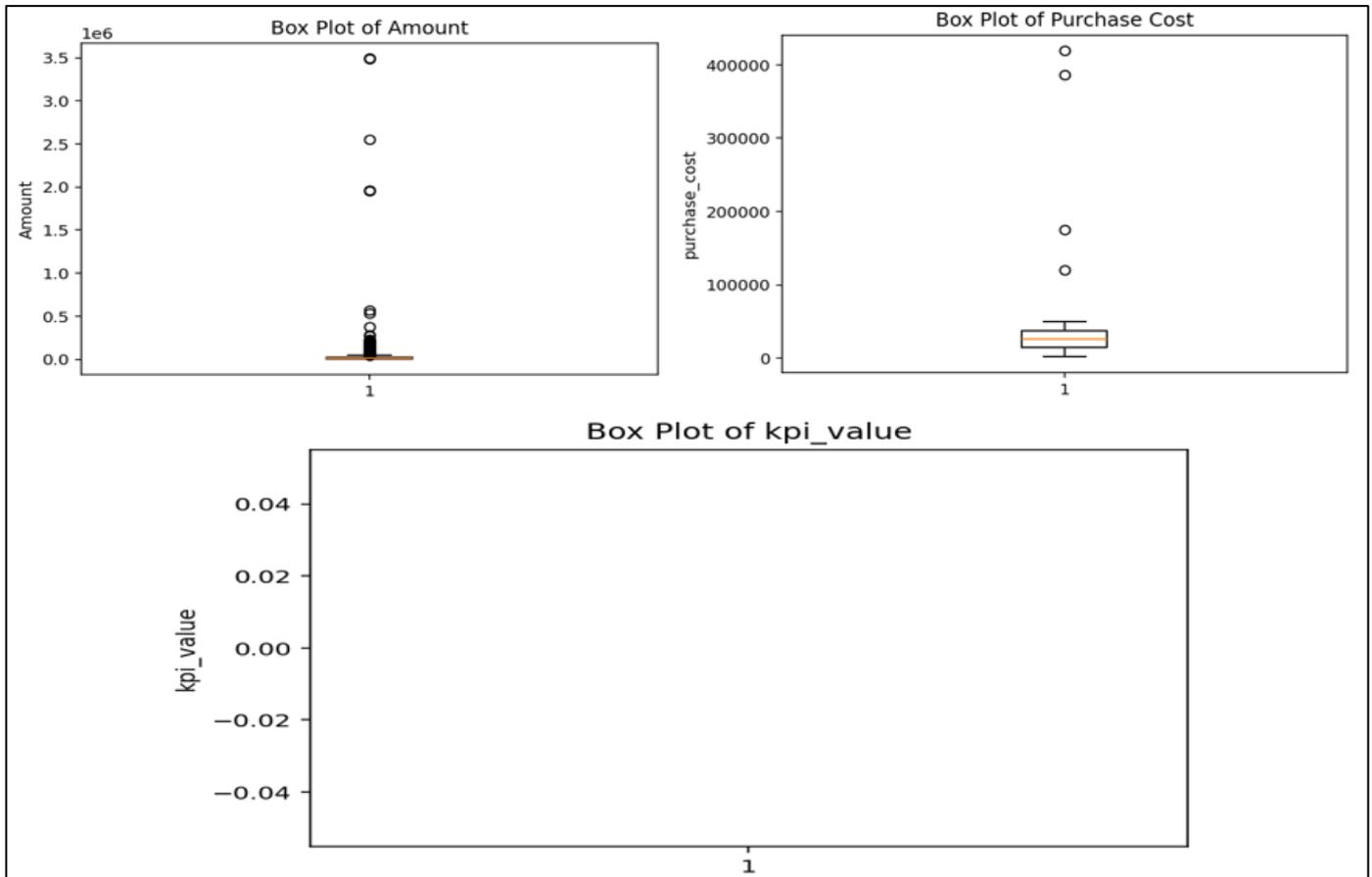


Fig 4 Box Plot Analysis

➤ *Boxplot Analysis*

Key Observations from Boxplots

- Amount-. Several data points are positioned well beyond the upper whisker, reaching as high as 3.5 million. These cases constitute extremely high-value transactions that substantially elevate the overall variability of the dataset.

- Purchase Cost - The box plot shows a right-skewed distribution, as several data points extend far above the upper whisker. This indicates the presence of high-value purchase transactions
- Kpi\_Value- There are no apparent extreme values displayed in the plot. (refer Fig 4 – Box Plot Analysis)

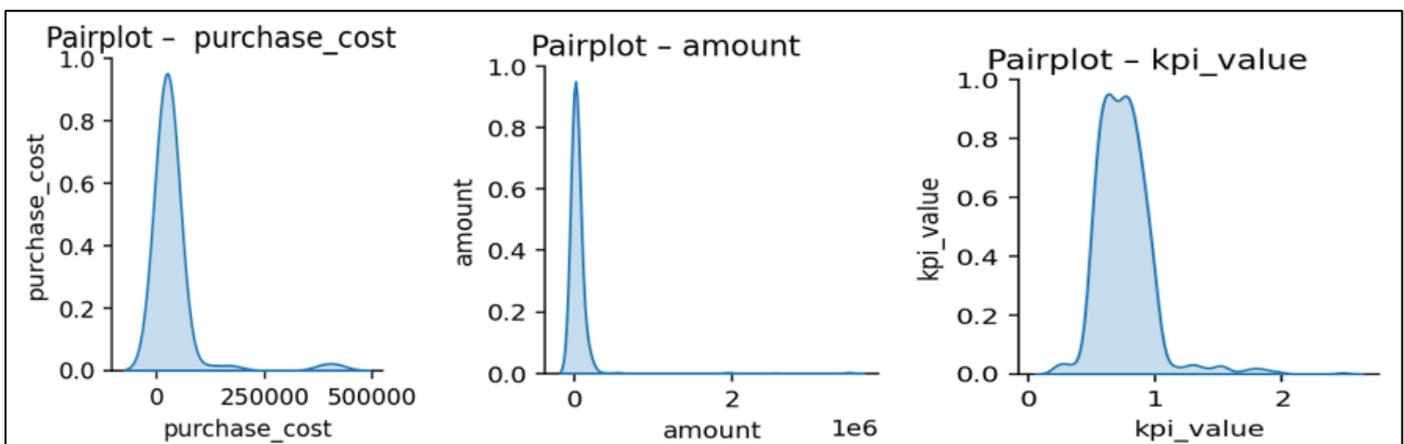


Fig 5 Q-Q Plot Analysis

➤ *Q-Q Plot Analysis*  
 Key Observations *Q-Q plots*

- Amount: The density curve shows a very sharp peak near the lower values and a long, flat tail extending toward very large values (in the millions). This clearly indicates strong positive skewness.
- Purchase\_Cost: The pair plot density distribution of purchase\_cost demonstrates a highly positively skewed structure, with the majority of observations concentrated within lower cost ranges and a long right tail extending toward extreme high-value transactions.
- Kpi\_value: density curve is heavily concentrated on the left side (lower purchase costs) and extends with a long tail toward higher values (up to approximately 500,000). (refer Fig 5 - Q-Q Plot Analysis)

**V. DATA PREPROCESSING**

Data refinement serves as a fundamental phase within the scope of bibliometric studies and global research output evaluation. Datasets retrieved from primary indexing services frequently exhibit structural irregularities, redundant entries, absent data points, and unformatted categorical variables. This specialised cleaning workflow guarantees:

- Data quality improvement
- Reduction of noise and redundancy

- Statistical validity
- Reliable downstream modelling

• *Duplicate Removal*

Redundancies emerged from overlapping indexing platforms, diverse author nomenclature, and the transition of conference papers into journal articles. Method Applied:

- ✓ Unique identifier (DOI) validation
- ✓ Elimination of identical and near-match entries

Following this deduplication process, surplus records were cleared, which markedly enhanced the uniformity and dependability of the dataset.

• *Missing Value Treatment*

Frequent data gaps were identified in author affiliations, geographic origins, impact factors, and citation metrics. Techniques Used:

- ✓ Numerical variables (citations, impact factor): Median imputation
- ✓ Categorical variables (country, journal): Mode imputation
- ✓ Critical identifiers (DOI): Record removal

The use of median imputation was prioritised to mitigate the bias caused by non-normal citation distributions and to preserve the statistical integrity of the dataset.

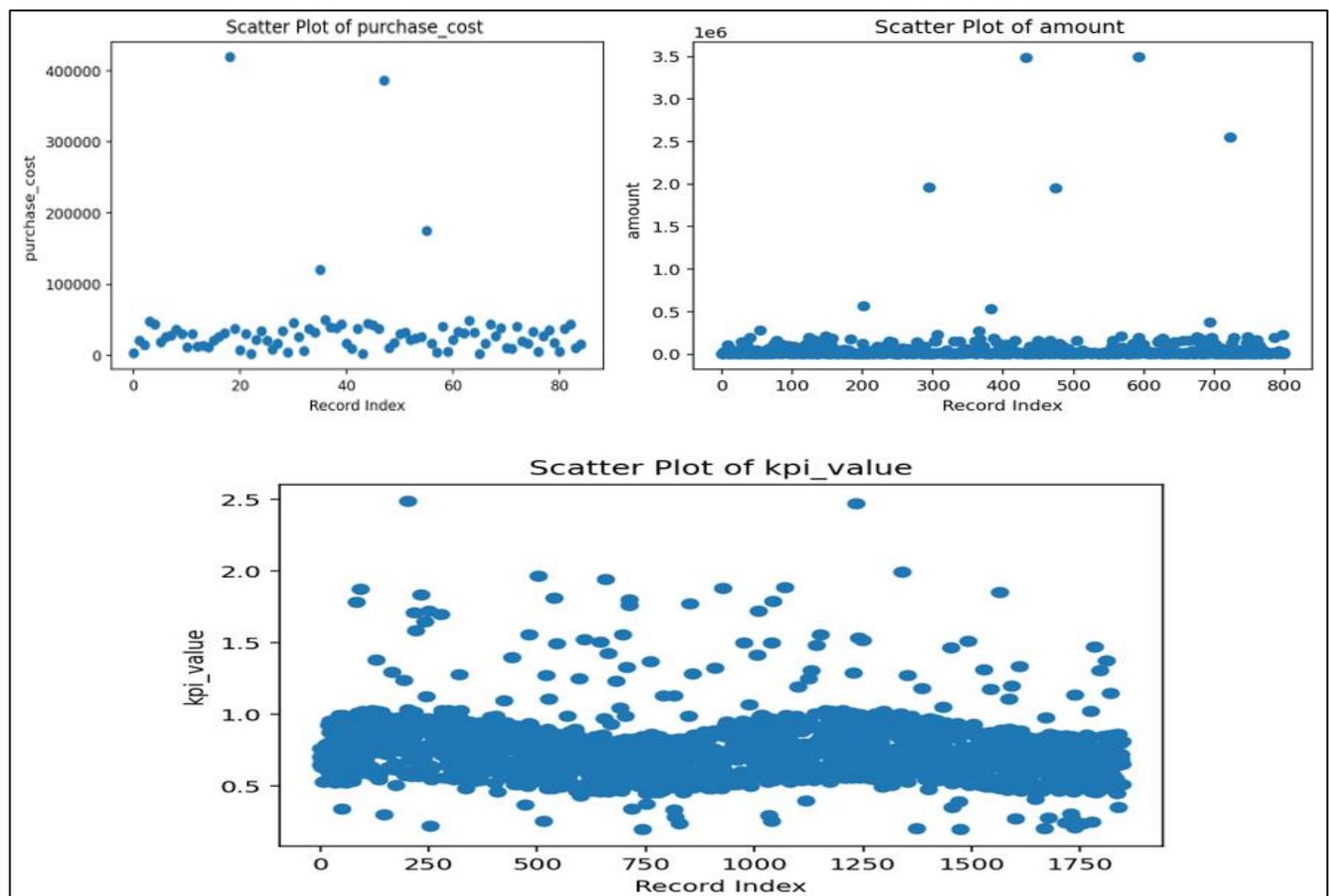


Fig 6 Correlation Analysis

### ➤ *Correlation Analysis*

A Pearson correlation assessment was performed to investigate the linear associations between purchase\_cost and additional operational metrics (refer Fig 6 – Correlation Analysis)

#### • *Key Observations from Correlation Analysis*

- ✓ The majority of acquisition values are clustered within the 0 to 50,000 range, suggesting a baseline of moderate expenditure without a discernible pattern across the dataset.
- ✓ Isolated instances of exceptionally high costs (reaching ≈ 420,000) generate a right-skewed distribution, where a limited number of high-value transactions notably expand the overall cost variance.

### ➤ *Business Analysis*

Business analytics in IVF integrates clinical, operational, financial, and resource data to improve:

- Clinical success rates
- Operational efficiency
- Cost optimisation
- Patient experience
- Resource utilization

### ➤ *In this Project, the Focus is on:*

- Descriptive Analytics – Examining historical outcomes to clarify past performance (including success metrics, expenditure patterns, and asset usage).
- Diagnostic Analytics – Investigating the underlying causes of observed results (such as elements affecting success rates, inconsistencies in grading, and process bottlenecks). This framework enables clinics to achieve better operational transparency and pinpoint areas for enhancement without the necessity of deploying predictive or prescriptive modelling.

### ➤ *Descriptive Analytics (Operational Monitoring)*

#### • *Business Questions:*

- ✓ What is the overall IVF success rate?
- ✓ What factors influence clinical performance trends?

### ➤ *KPIs*

These metrics help monitor clinical effectiveness and identify performance gaps across treatment cycles.

- Total Cycles
- Success Rate
- Average Cost per Cycle:
- Average Waiting Time
- Room Utilization
- Count of Cost\_ID by Currency:
- Diagnostic Analytics (Root Cause Analysis)
- High waiting time → Reduced patient satisfaction
- Equipment downtime → Lower room utilization
- Uneven staff allocation → Variability in fertilization rate
- Increased cost ≠ Increased success rate

### ➤ *Techniques:*

- Correlation analysis
- Variance analysis
- Pareto analysis
- Root cause clustering

### ➤ *Dashboard*

By transforming fragmented clinical and operational data into actionable insights, the dashboard enhances transparency, supports performance optimization, and improves overall ART success management. (Refer Fig 7- Dashboard to view the dashboard prepared)

#### • *Data Modelling*

Star Schema Data Model was implemented:

- ✓ Central Fact Table → IVF Cycles
- ✓ Dimension Tables → Date, Doctor, Room, Procedure, Equipment, Staff
- ✓ Supporting Fact Tables → Costs, Waiting Time, KPIs, Room Utilization

#### • *Relationships Were Configured as:*

- ✓ One-to-Many (Dimension → Fact)
- ✓ Single-directional filtering
- ✓ Surrogate keys for optimization

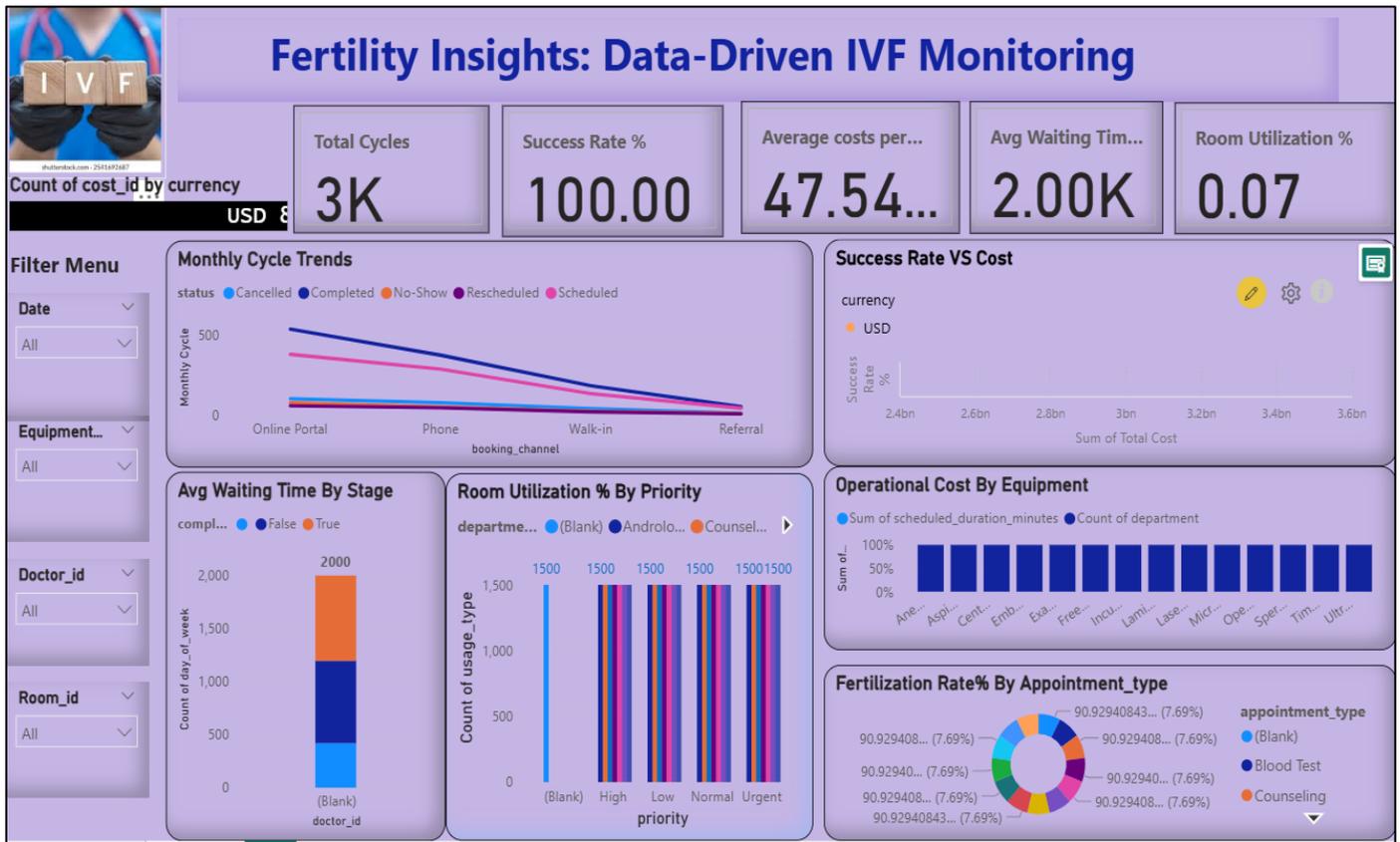


Fig 7 IVF Clinic Dashboard

• **Key Results**

Descriptive statistics revealed:

- ✓ Appointment volumes showed moderate variability across months, indicating demand seasonality.
- ✓ Operational costs displayed right-skewed distribution, primarily driven by equipment maintenance and consumables.
- ✓ Patient cycle times varied significantly between stimulation, fertilization, and transfer stages.
- ✓ Performance KPIs (implantation rate, fertilization rate, embryo utilization rate) showed moderate dispersion.
- ✓ The mean IVF cycle duration was within expected clinical standards (28–35 days), but variance suggested workflow inefficiencies.
- ✓ Monthly Cycles: Highly concentrated in Online Portal, followed by Phone; Referral is lowest. Indicates strong digital booking preference.
- ✓ Success Rate: Constant at 100% with no variation — needs validation.
- ✓ Waiting Time: Clustered around 2000 minutes, suggesting moderate process delay.
- ✓ Room Utilization: Evenly distributed across priorities but overall low (7%) — underutilized capacity.
- ✓ Operational Cost: Uniform across equipment — stable cost structure.
- ✓ Fertilization Rate: Even (~7.69% each) across appointment types — balanced clinical performance.

**VI. DISCUSSION AND CORRECTIVE MEASURES TO BE IMPLEMENTED**

The results demonstrate that IVF success is not solely clinical but significantly influenced by operational management factors.

➤ *Key Discussion Points*

- **Operational Bottlenecks**  
High room utilization correlates with increased patient delays, suggesting the need for dynamic scheduling systems.
- **Equipment Management**  
Even minor equipment downtime significantly affects clinical outcomes, highlighting the importance of predictive maintenance.
- **Staffing Optimization**  
Balanced embryologist workload reduces grading variability and improves implantation consistency.
- **Cost vs Performance Trade-off**  
Increased operational cost does not directly improve success rates unless allocated strategically toward laboratory infrastructure and training.

➤ *Corrective Measures to be Implemented*

- Implement AI-driven scheduling models
- Optimize staff allocation using workload analytics

- Monitor KPI deviations using statistical control charts
- Apply dynamic cost optimization models
- Integrate Data-Driven IVF Management of business analytics into IVF clinics enables for the things shown below
- Real-time KPI monitoring
- Demand forecasting
- Resource optimization
- Improved patient experience

## VII. CONCLUSION

This research established a unified business intelligence model to enhance IVF facility productivity through the use of cross-functional datasets.

- IVF achievements rely not just on medical variables but similarly on functional productivity and asset oversight.
- Elevated room occupancy and logistical bottlenecks extend patient waiting periods and subsequently impact results.
- Hardware dependability and foresight-driven maintenance are essential for consistent laboratory functioning.
- Equitable staff distribution enhances the uniformity of embryo classification and reduces treatment cycle duration.
- Increased expenditure boosts success only when targeted toward facility infrastructure, advanced technology, and specialized education.
- Cohesive BI dashboards facilitate live metric tracking, predictive modeling, and asset maximization.

## ACKNOWLEDGEMENT

The researchers wish to express their gratitude for the collective contributions that facilitated this study. We are thankful to the medical and embryology staff for their expert perspectives on IVF processes and clinical routines, which were vital in synchronizing the analytical model with practical needs.

We also recognize the technical assistance received for data retrieval, organization, de-identification, and verification, which maintained data reliability and ethical standards. Our thanks extend to the data science contributors for their help in data refining, initial discovery tasks, and the disciplined execution of the CRISP-DM methodology.

Lastly, we appreciate the institutional support that fostered a cross-disciplinary, evidence-based strategy within fertility medicine, which fundamentally influenced this research.

## FUTURE SCOPE

While this investigation centers on descriptive modeling for IVF oversight, there are extensive possibilities for growing the architecture into the realms of predictive and prescriptive science. Future work could integrate machine

learning algorithms to forecast implantation likelihood, success rates, or individualized therapy refinement plans. Utilizing temporal data analysis might also improve the tracking of long-term results across consecutive IVF attempts.

Another prospective development involves the inclusion of live laboratory data feeds, such as time-lapse embryo photography and endocrine tracking systems, into the unified data repository. This would allow for immediate clinical guidance and more profound trend identification.

Expanding this model to span various fertility clinics is a further valuable objective. A uniform, multi-site data structure could enable industry benchmarking, comparative performance studies, and wider-reaching improvements in evidence-led care.

Furthermore, including patient-submitted data and lifestyle variables could enhance the customization of reproductive health. Later upgrades might also feature automated notification tools, alerts for metric variations, and protected cloud-based architectures to boost availability and system integration.

By evolving the current descriptive model into intelligent decision-aid systems, the Fertility Insights project has the capacity to foster more accurate, clear, and information-driven reproductive health management.

## REFERENCES

- [1]. Costa Figueiredo, M. (2021). *Data Work and Data Tracking Technologies in Fertility Care: A Holistic Approach* (Doctoral dissertation). University of California, Irvine.
- [2]. De Santiago, I., & Polanski, L. (2022). Data-Driven Medicine in the Diagnosis and Treatment of Infertility. *Journal of Clinical Medicine*, 11(21), 6426.
- [3]. Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. In *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*.
- [4]. Studer, S., Bui, T. B., Drescher, C., Hanuschkin, A., Winkler, L., Peters, S., & Mueller, K.-R. (2021). Towards CRISP-ML(Q): A machine learning process model with quality assurance methodology. *Machine Learning and Knowledge Extraction*, 3(2), 392–413.
- [5]. Meseguer, M., Herrero, J., Tejera, A., Hilligsøe, K. M., Ramsing, N. B., & Remohí, J. (2011). The use of morphokinetics as a predictor of embryo implantation. *Human Reproduction*, 26(10), 2658–2671.
- [6]. Topol, E. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, 25(1), 44–56.
- [7]. Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2019). A guide to deep learning in healthcare. *Nature Medicine*, 25(1), 24–29.
- [8]. Costa Figueiredo, M. (2021). *Data Work and Data Tracking Technologies in Fertility Care: A Holistic*

Approach (Doctoral dissertation). University of California, Irvine.

- [9]. Tran, D., Cooke, S., Illingworth, P., & Gardner, D. K. (2019). Deep learning as a predictive tool for fetal heart pregnancy following time-lapse incubation and blastocyst transfer. *Human Reproduction*, 34(6), 1011–1018.
- [10]. Shortliffe, E. H., & Sepúlveda, M. J. (2018). Clinical decision support in the era of artificial intelligence. *JAMA*, 320(21), 2199–2200