Analysis of IVF Trends in the United States (2020 - 2022)

Svitlana Musiienko

Learner of MIT Emerging Talent Program in Data Science
Cohort 6, ELO2

Why This Project Matters:

- 1 in 6 couples experience infertility;
- IVF treatment is emotionally difficult and success is not guaranteed, it varies widely by age and treatment approach;
- my personal IVF experience insights me that success depends on many factors that patient rarely see clearly;
- understanding these trends supports better patient counselling, clinic transparency and public health planning.

Data Source

All data downloaded from **CDC National ART Surveillance System (NASS)**

Years:

2020 - 2022

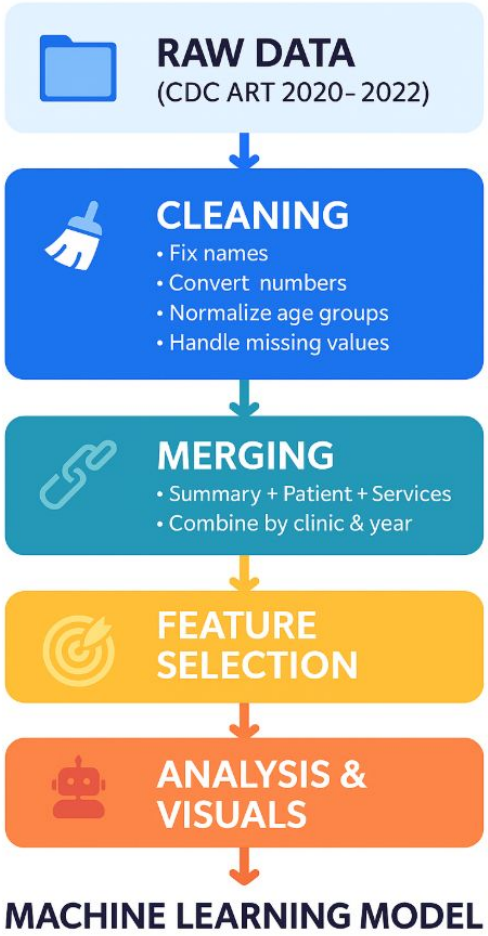
Reports used for analysis of each year (12 files):

- ❏ Assisted Reproductive Technology Patient and Cycle Characteristics;
- ❏ Assisted Reproductive Technology Services and Profiles;
- ❏ Assisted Reproductive Technology Summary.

The datasets include patient age, clinic services, cycle counts, retrievals, transfers, donor usage, frozen cycles, gestational carrier use, lab accreditation, and detailed success rate indicators (live birth %, singleton %, embryo transfer %, cancellation %, etc.)

Data Pipeline

turning raw Assisted Reproductive Technology (ART) data into usable insights



Analysis Highlights

Analysis and visualizations shows us the story:

“ Where you live, how busy your clinic is and your age all meaningfully shape IVF success ”

🗺️ **U.S. IVF Clinic Distribution Map -** states like **California, Texas, New York, Florida** have the highest concentrations of clinics , many central and rural states have very limited access. Patients in low-density states likely face travel, cost burdens, and delayed care.

📈 IVF Cycle Volume (2020–2022) - treatment volume **dropped in 2020** (pandemic shutdowns), It then **surged in 2021** and stabilized in 2022. Cycle volume correlates with clinic expertise and resources.

📉 Age-Based Live-Birth Success Curves - success rates drop **sharply after age 37**, patients aged **<35** have nearly double the success rates compared to women **38–40**, outcomes for women **>40** are significantly lower across all years.

ML Modeling

Two complementary machine learning models were built to understand what predicts high IVF success at the clinic level

Features for ML models:

Age group distribution - *patients <35, 35–37, 38–40, >40*

Clinic volumec - *number of retrievals/transfers performed*

State - *capturing geographic and policy differences*

📊 Logistic Regression Model

Identify which factors increase or decrease the probability of high success.

Insights:

- younger age groups (<35) strongly increase the odds of high success;
- patients >40 strongly decrease the odds;
- clinic volume is positively associated with better outcomes;
- some states show systematic differences.

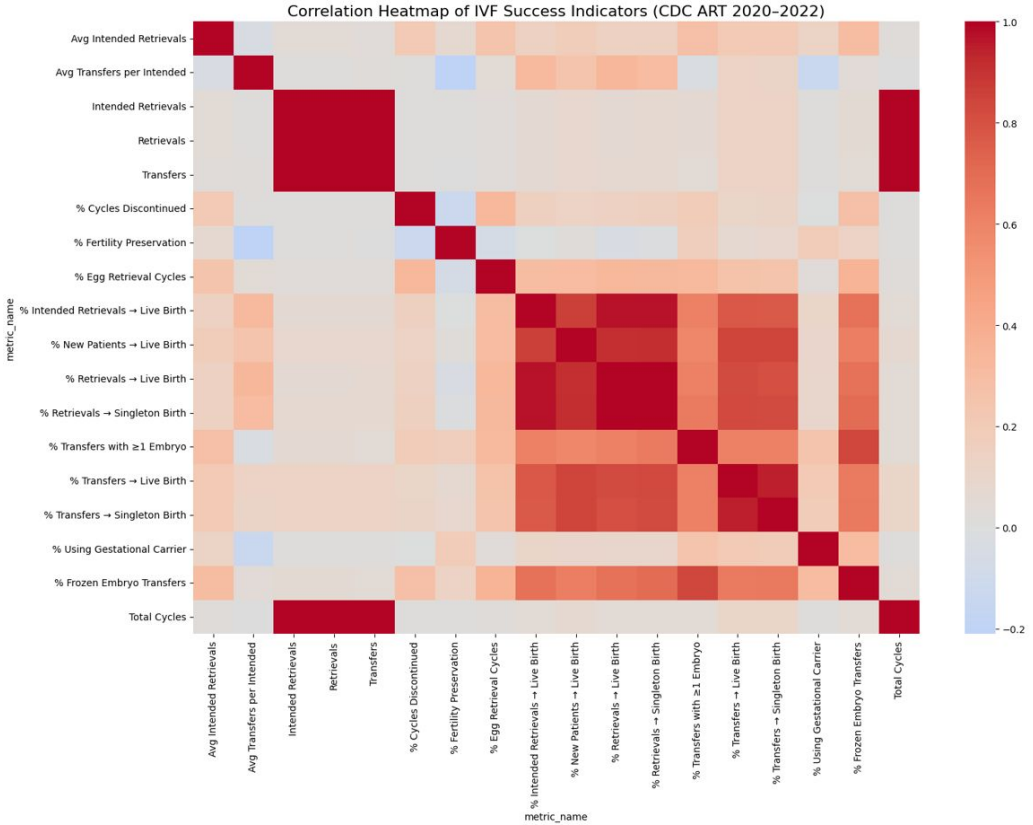
🌳 Decision Tree

Understand decision pathways and thresholds.

Insights:

- clinic volume emerges as the top split → high-volume clinics outperform;
- age groups form clear branches, confirming age as a decisive factor;
- state remains a weaker but meaningful contributor.

Correlation Heatmap of IVF Success Indicators



🔴 1. Strong, Positive Clusters (Dark Red Zones):

- Retrieval success → Transfer success → Live birth success;
- Singleton live births closely track overall live-birth success;
- Transfers per retrieval connect strongly to total success metrics.

🔵 2. Moderate Correlations:

- Frozen embryo transfers moderately correlate with higher success;
- Intended vs actual retrieval metrics often move together;
- Cycle volume shows weak-to-moderate association.

⚪ 3. Weak or Negative Correlations (Light Colors):

- Fertility preservation cycles have almost no relationship to success metrics;
- Cycle discontinuation rates sometimes trend negatively.

Final Summary

- 🧬 **Age Matters - Significantly**
  Success rates decline sharply after 37, with the largest drop after age 40.
- 🏥 **Clinic Volume Predicts Better Outcomes**
  High-volume clinics consistently achieve higher live-birth rates and fewer discontinued cycles.
- 📍 **Geography Shapes Access & Success**
  Large differences exist between states in both clinic availability and performance.
- 🔗 **Indicators Work Together**
  Retrieval success, transfer success, and live-birth metrics form a strong, correlated pathway.
- 🤖 **ML Confirms the Patterns**
  Logistic Regression & Decision Trees identify age group, clinic volume, and state as key predictors of above-average success.

Impact for the Public:

These insights make IVF more understandable and accessible. By revealing how age, clinic volume and geography shape success rates, the public gains clearer expectations, more informed decision-making power, and better awareness of disparities in reproductive healthcare. This supports individuals and families on their fertility journeys and helps drive more equitable access to treatment across the country.

THANK YOU!

Svitlana Musiienko

MIT Emerging Talent – Data Science Program
Project: *Analysis of IVF Trends in the United States (2020-2022)*

📍 Focus: Women's health • IVF outcomes • Data-driven insights
📧 Email: lanamus88@gmail.com

🔗 **GitHub project:**
https://github.com/lanamusienko/fertility-trends

Scan to view the full project: