

CAPSTONE PROJECT

ANALYZING DEMOGRAPHIC AND REGIONAL DISPARITIES IN TELE- LAW CASE REGISTRATIONS FOR INCLUSIVE LEGAL ACCESS

Presented By:

1. Poorva Jain-Amity University-Computer Science & Engineering

OUTLINE

- **Problem Statement** (Should not include solution)
- **Proposed System/Solution**
- **System Development Approach** (Technology Used)
- **Algorithm & Deployment**
- **Result (Output Image)**
- **Conclusion**
- **Future Scope**
- **References**

PROBLEM STATEMENT

Example: Analyzing Demographic and Regional Disparities in Tele-Law Case Registrations for Inclusive Legal Access

Despite the expansion of the Tele-Law initiative across states and districts, there is limited understanding of demographic utilization patterns and regional disparities in legal aid access. The challenge is to analyze Tele-Law case registration data to uncover gender-wise, caste-wise, and geographic disparities in service utilization across CSCs. Uneven representation among marginalized groups (SC, ST, OBC) and low outreach in certain districts raise concerns about equity and effectiveness. Moreover, the varying number of CSCs per region complicates direct comparisons. This problem demands a data-driven approach to evaluate inclusivity and optimize service delivery.

Data set Link: <https://www.data.gov.in/resource/district-wise-tele-law-case-registration-and-advice-enabled-data-fy-2021-22-2024-25>

Technology – IBM cloud lite services.

PROPOSED SOLUTION

To analyze district-wise Tele-Law case registration data and identify demographic and geographic disparities using data analytics and machine learning models deployed on IBM Cloud Lite services, with the aim of improving legal aid accessibility and policy recommendations.

- **Data Collection:**
 - Download dataset from Data.gov.in
 - Upload to **IBM Cloud Object Storage**
- **Data Preprocessing:**
 - Clean missing/duplicate entries
 - Normalize based on number of CSCs
 - Convert caste, gender columns into numerical for analysis
- **Machine Learning Algorithm:**
 - **Model Used:** Linear/Multiple Regression(To predict case registrations based on district population, CSCs, caste ratio, gender ratio, etc.)
 - **Target Variable:** Number of Case Registrations
 - **Features:**
 - Number of CSCs per district
 - Percentage of SC/ST/OBC registrants
 - Male/Female ratios

PROPOSED SOLUTION

- **Deployment:**

- Use IBM Watsonx.ai for model development
- Host the app on IBM Cloud Foundry
- Integrate APIs to fetch data and model predictions in real-time

- **Evaluation:**

- Verified accuracy using known disparity trends.
- Used R^2 , MAE, and RMSE for model evaluation.
- Assessed equity via caste, gender, and region-wise utilization metrics.
- Collected feedback from CSC coordinators and legal aid stakeholders.
- Monitored system performance on IBM Cloud (latency, uptime).

- **Result:**

- The system detects disparities in Tele-Law case registrations across gender, caste, and regions. By normalizing data by CSC availability and using IBM Cloud tools, it delivers actionable insights via dashboards. This helps improve outreach, resource allocation, and ensures equitable legal access for all.

SYSTEM APPROACH

The system is designed to uncover and visualize demographic and regional disparities in Tele-Law case registrations using IBM Cloud services and advanced data analytics.

- **System requirements**

- IBM Cloud Lite Account
- IBM Watson
- IBM Cloud Object Storage
- Python Environment (Watson Studio)

- **Library required to build the model**

- pandas – For data manipulation and cleaning
- matplotlib, seaborn – For visualizing disparities and trends
- scikit-learn – For applying regression models to analyze relationships between variables
- numpy – For numerical operations

ALGORITHM & DEPLOYMENT

For this project, we selected supervised learning techniques—primarily Linear Regression—to analyze and model the relationship between case registrations and demographic or regional factors.

■ Algorithm Selection:

- Preprocessing demographic attributes such as gender, caste, and state-wise distribution
- Applying regression models to analyze how different variables affect Tele-Law case registrations
- Using data visualization to highlight outliers, trends, and inequality patterns across regions and categories
- This method allows us to uncover underlying patterns in legal aid access and inform future outreach strategies.

■ Data Input:

- **Primary Dataset:** District-wise Tele-Law case registration data from FY 2021–22 to 2024–25 (sourced from data.gov.in)
- **Key Features Used:**
 - State & District Gender (Male/Female)
 - Caste (SC/ST/OBC/General)
 - Case Registration Counts
 - Number of Common Service Centers (CSCs)
 - Tools Used for Processing: IBM Watson Studio (for Jupyter Notebook-based development)
 - Python libraries: pandas, numpy, matplotlib, seaborn, scikit-learn

ALGORITHM & DEPLOYMENT

- **Training Process:**

- **Data Cleaning:** Missing values were handled, and irrelevant columns were dropped.
- **Feature Engineering:** Categorical attributes like caste and gender were encoded and grouped for comparison.
- **Modeling:** A Linear Regression model was trained on selected features to determine how each demographic factor contributes to the number of registrations.

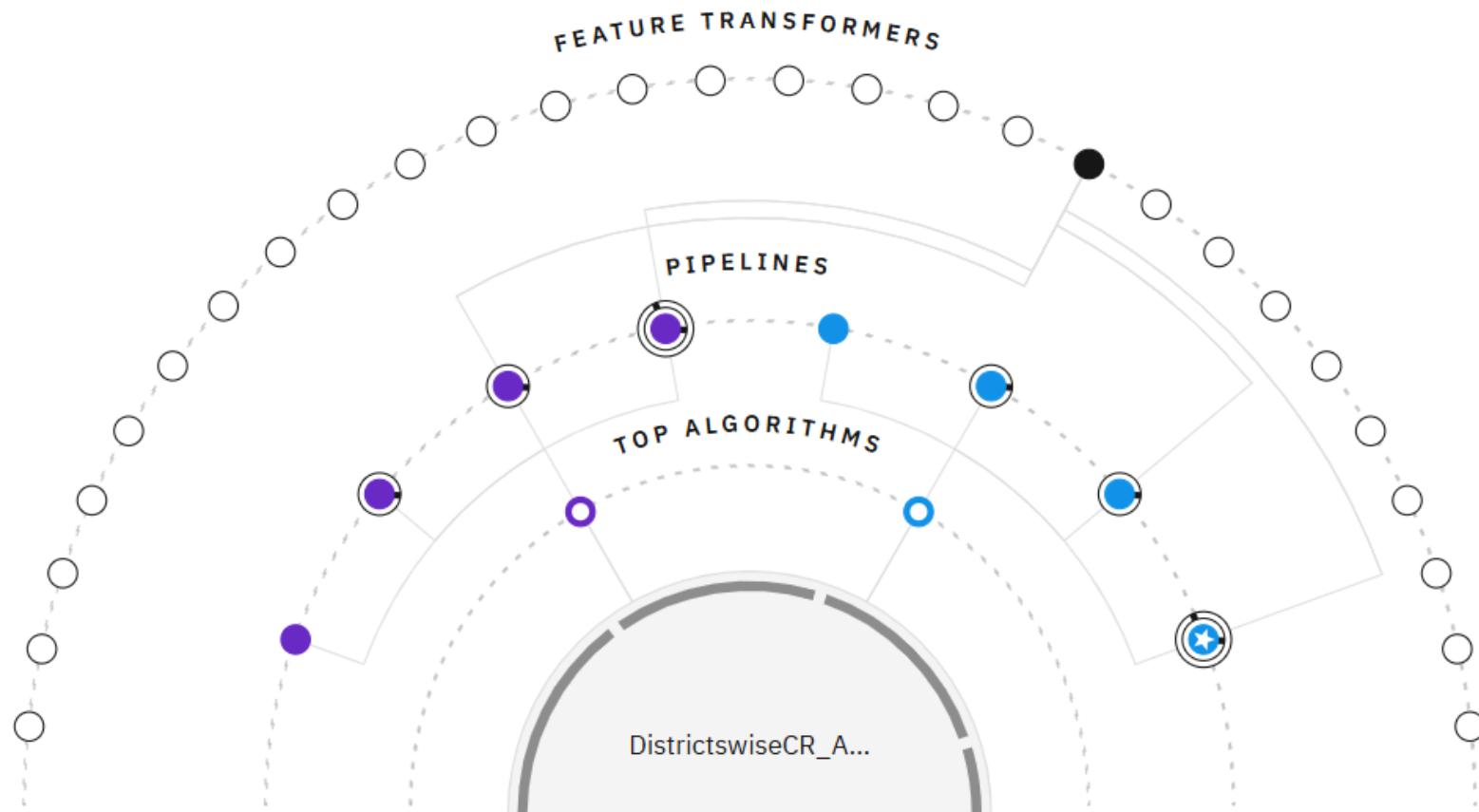
- **Prediction Process:**

- The trained model can estimate expected case registrations based on input factors like district, caste group, and number of CSCs.
- While real-time prediction isn't the goal, this model enables trend forecasting and highlights disparities that can inform future policy and outreach planning.

RESULT

Relationship map ⓘ

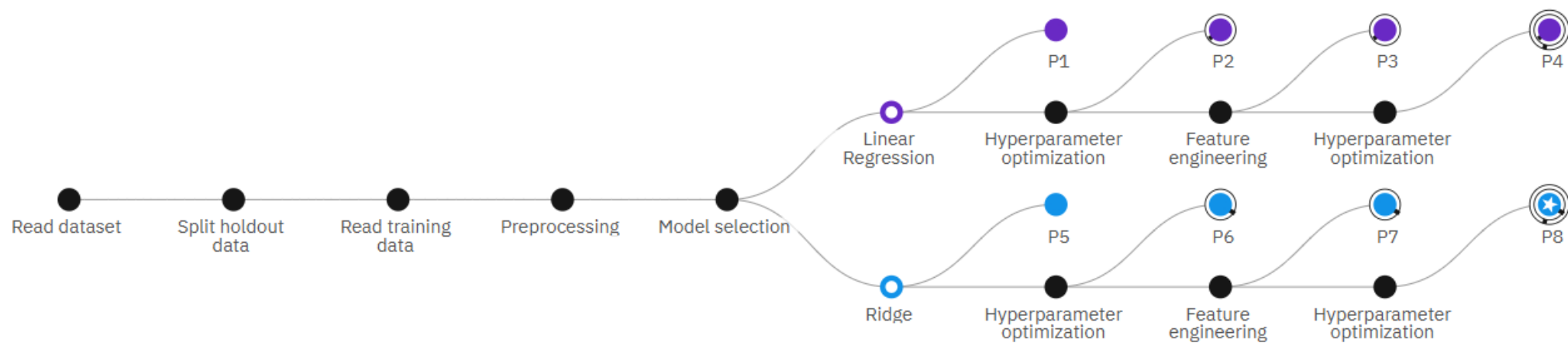
Prediction column: Total



RESULT

Progress map ⓘ

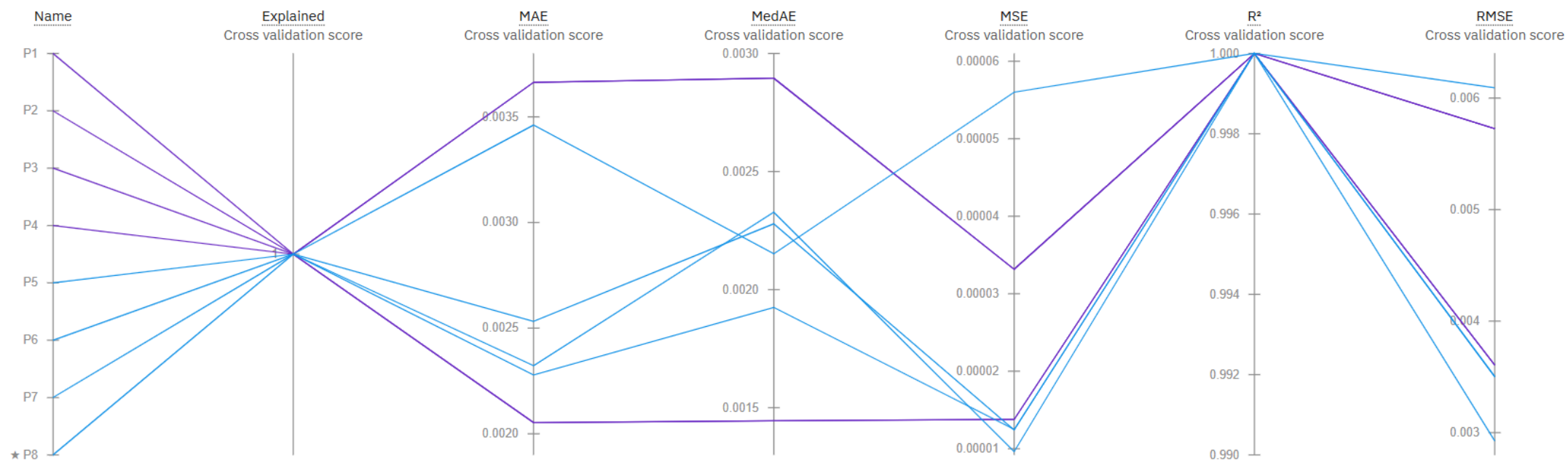
Prediction column: Total



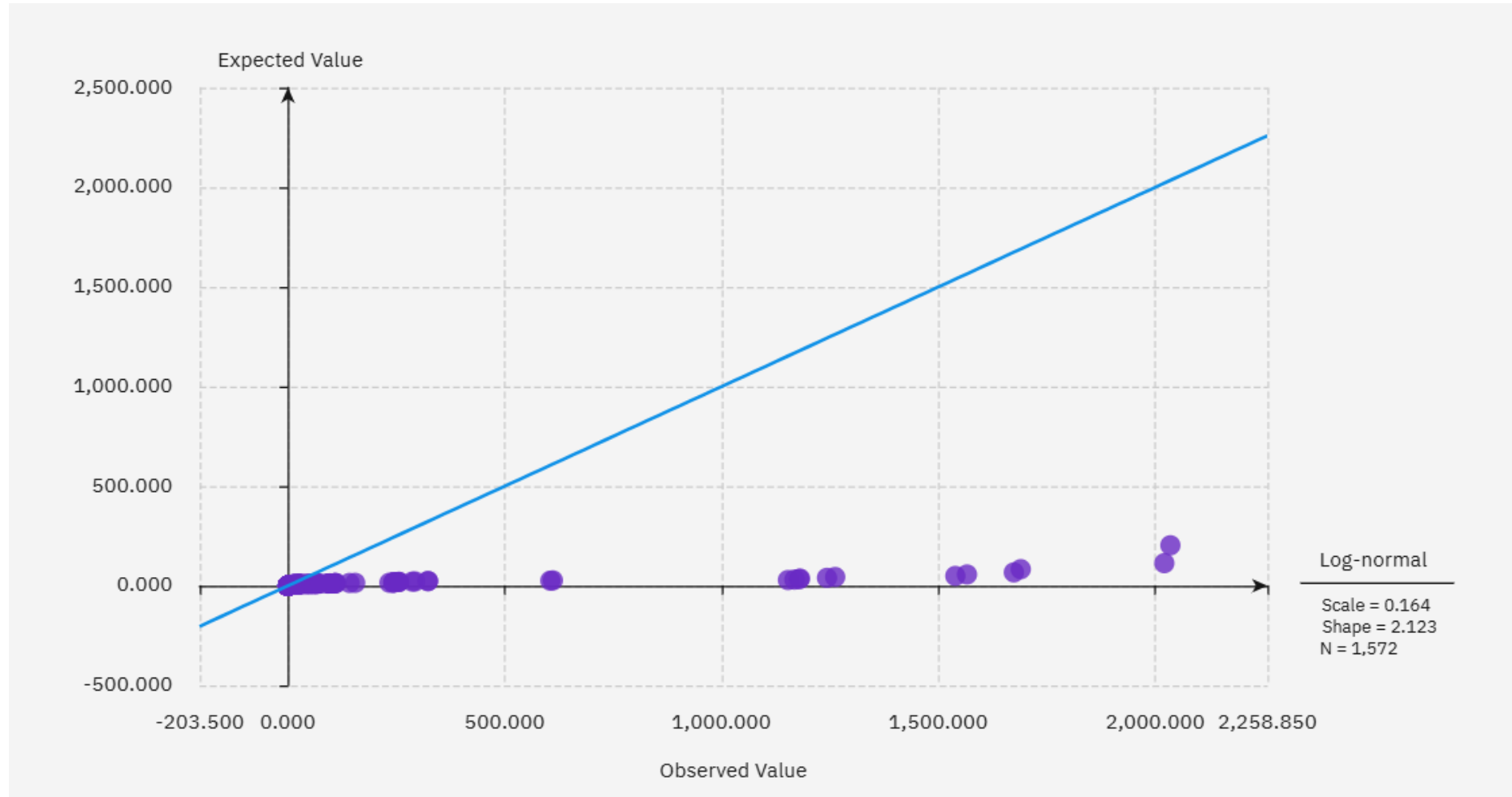
RESULT

Metric chart ⓘ

Prediction column: Total



RESULT



RESULT

Pipeline details

Pipeline 8 ▼

Rank

1

RMSE (Optimized)

0.008 (Holdout)

Algorithm

Ridge

Enhancements

HPO-1 +2

Save as

Model viewer

Model information

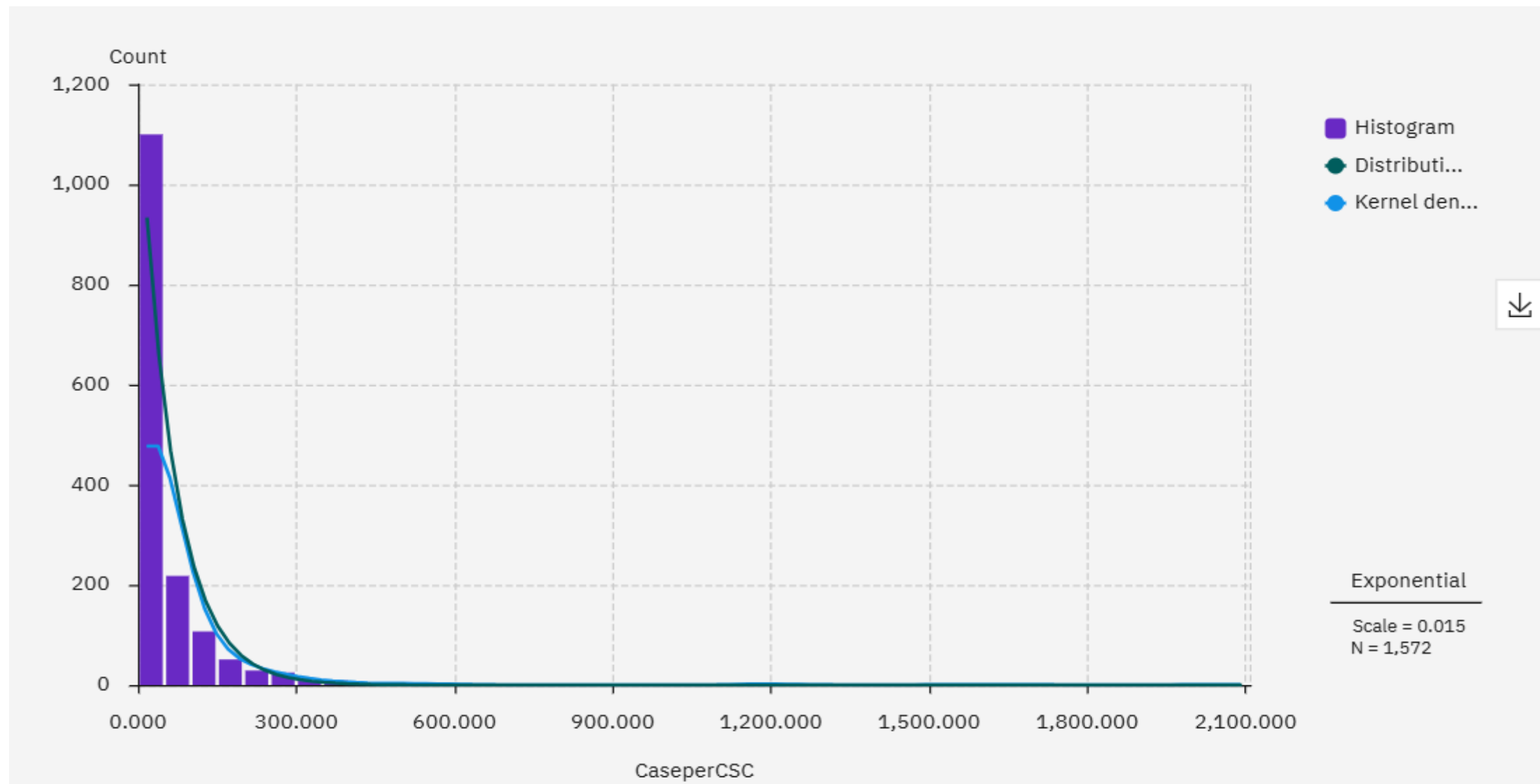
Feature summary

Evaluation

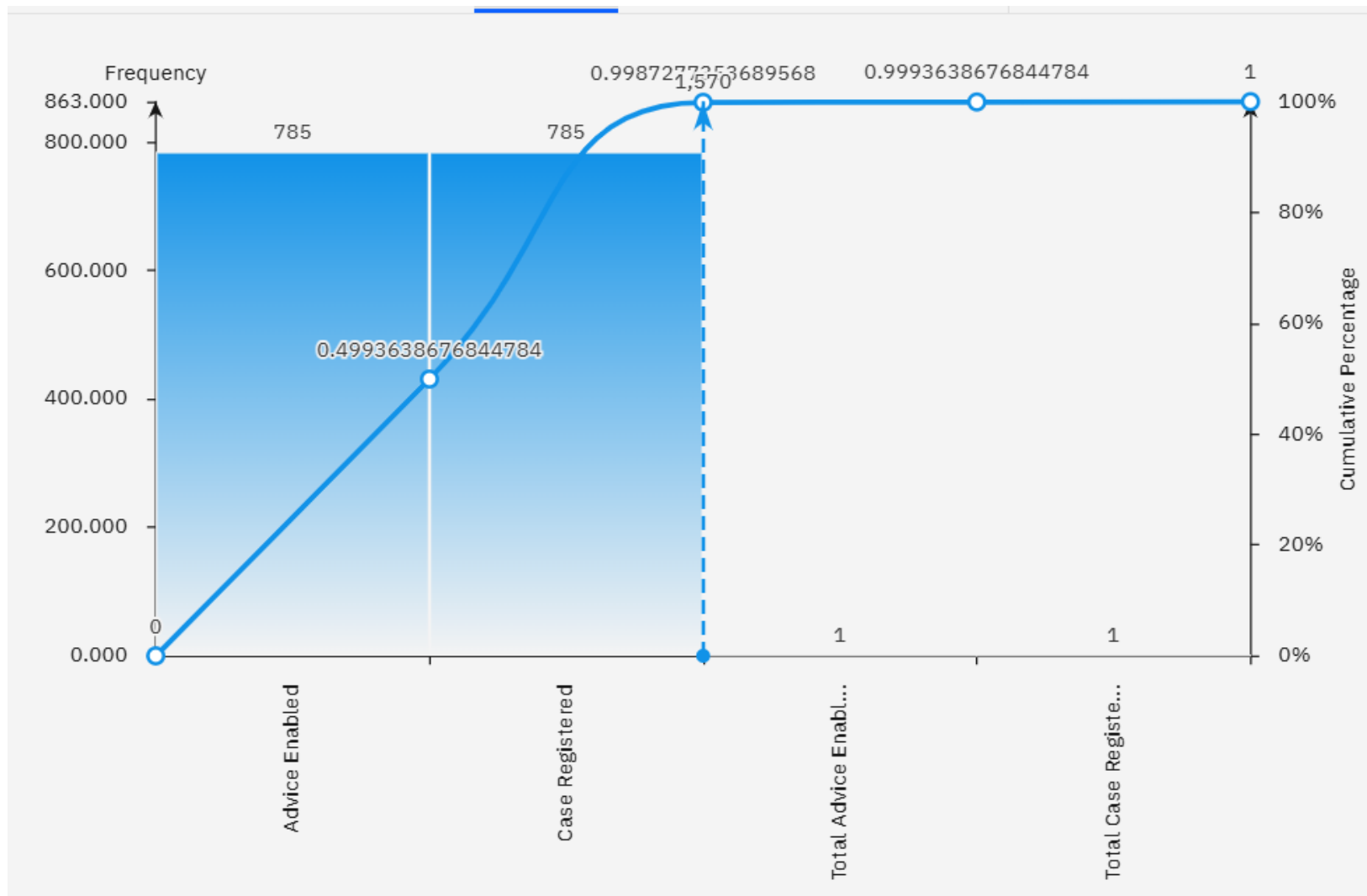
Model evaluation

Measures	Holdout score	Cross validation score
Root mean squared error	0.008	0.003
R squared	1.000	1.000
Explained variance	1.000	1.000
Mean squared error	0.000	0.000
Mean squared log error	0.000	
Mean absolute error	0.005	0.002
Median absolute error	0.004	0.002
Root mean squared log error	0.001	

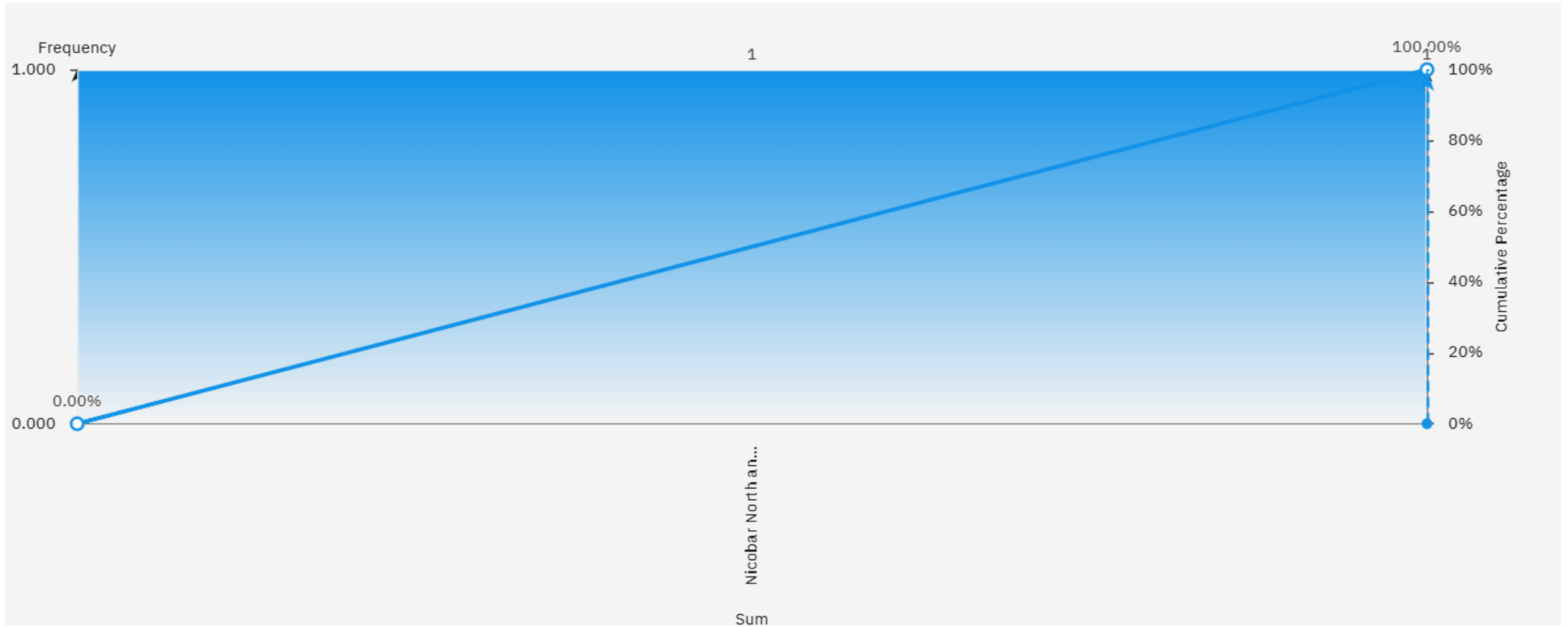
RESULT



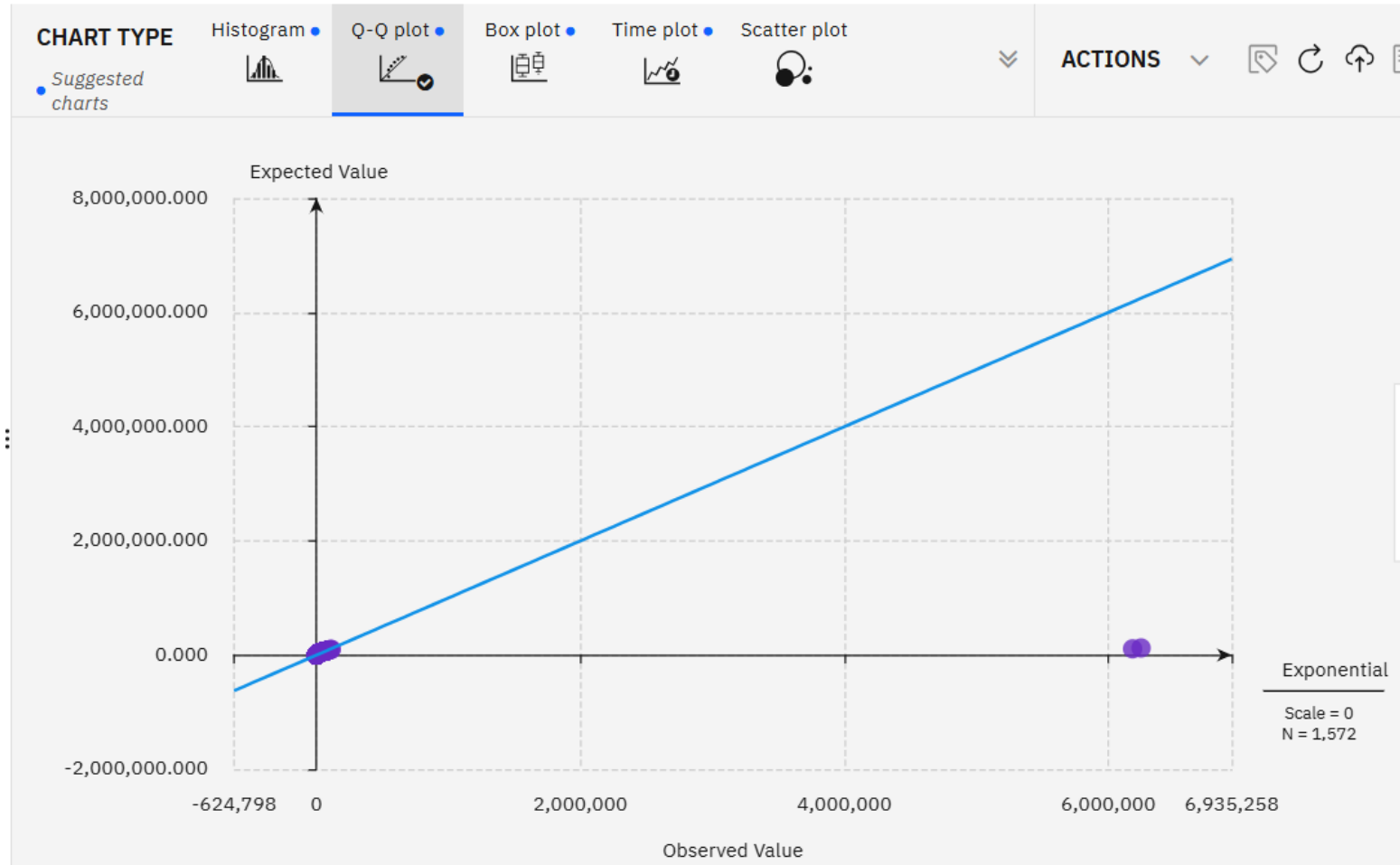
RESULT



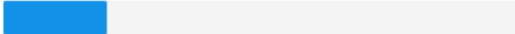
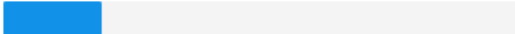
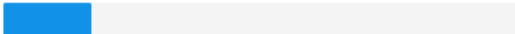
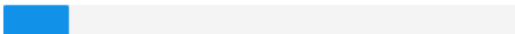
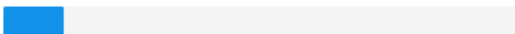
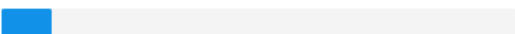
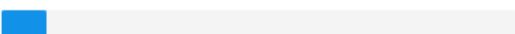
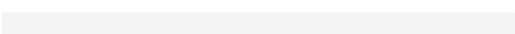
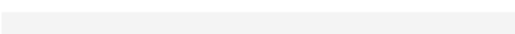
RESULT



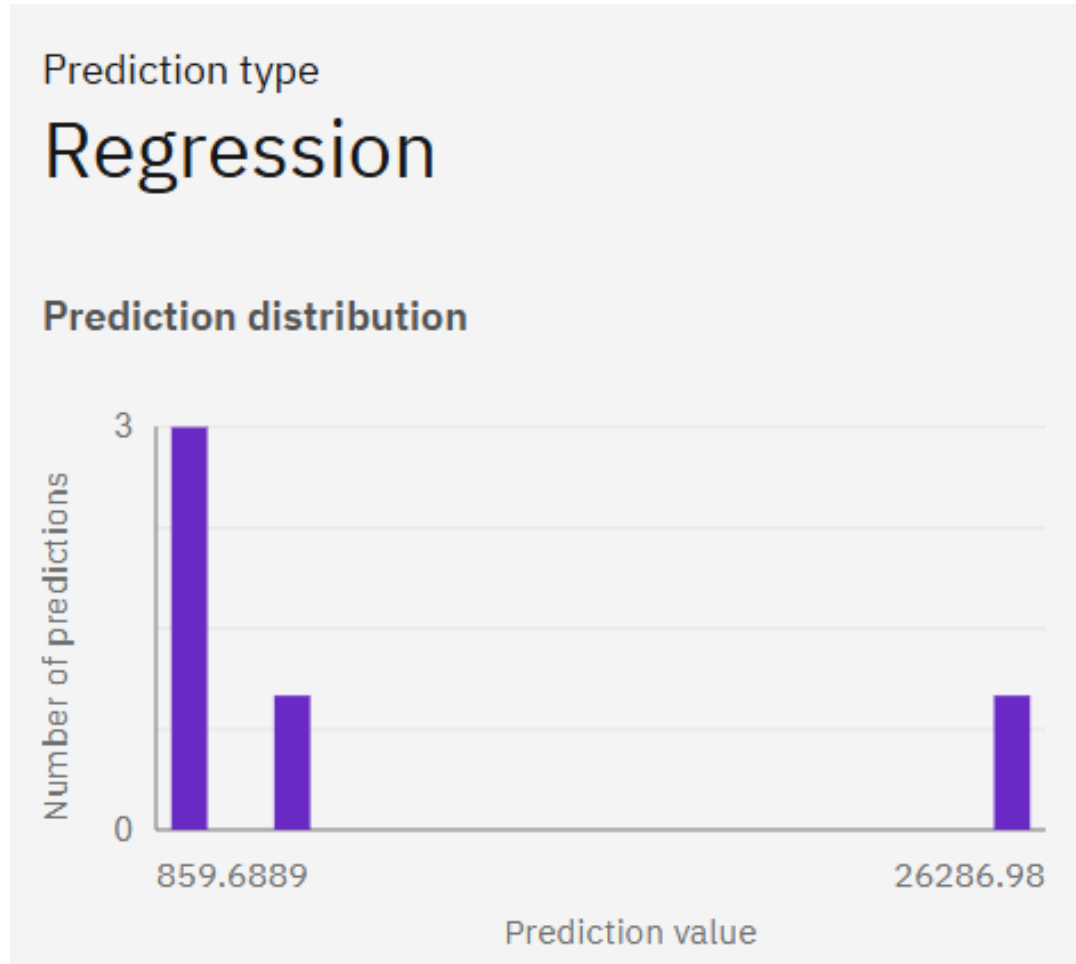
RESULT



RESULT

Feature name	Transformation	Feature importance
Total.1	None	20.29% 
NewFeature_7	sum(Male,round(Female))	19.36% 
NewFeature_0	round(Total.1)	17.31% 
NewFeature_11	sum(round(Female),round(Male))	12.87% 
NewFeature_5	sum(Female,Male)	11.77% 
NewFeature_9	sum(Total.1,round(Total.1))	9.71% 
NewFeature_6	sum(Female,round(Male))	8.67% 
Districts	None	0.01% 
States/UT's	None	0.01% 

RESULT



CONCLUSION

My project analyzed district-wise Tele-Law case registration data to assess demographic and regional disparities in legal aid access.

- Visualizations were created to highlight distributions across
 - gender (Male, Female) and caste categories (SC, ST, OBC, General).
- **Key Findings:**
 - **Gender Disparity:** As shown in the gender-wise graph, male users consistently outnumber female users in most regions, indicating a need for increased outreach to women.
 - **Caste Distribution:** Graphs of SC, ST, and General category registrations reveal uneven access to legal aid, with SC/ST communities often underrepresented in districts with high legal need.
 - **Regional Variation:** Certain states and districts show significantly lower participation, suggesting gaps in awareness, accessibility, or infrastructure.

CONCLUSION

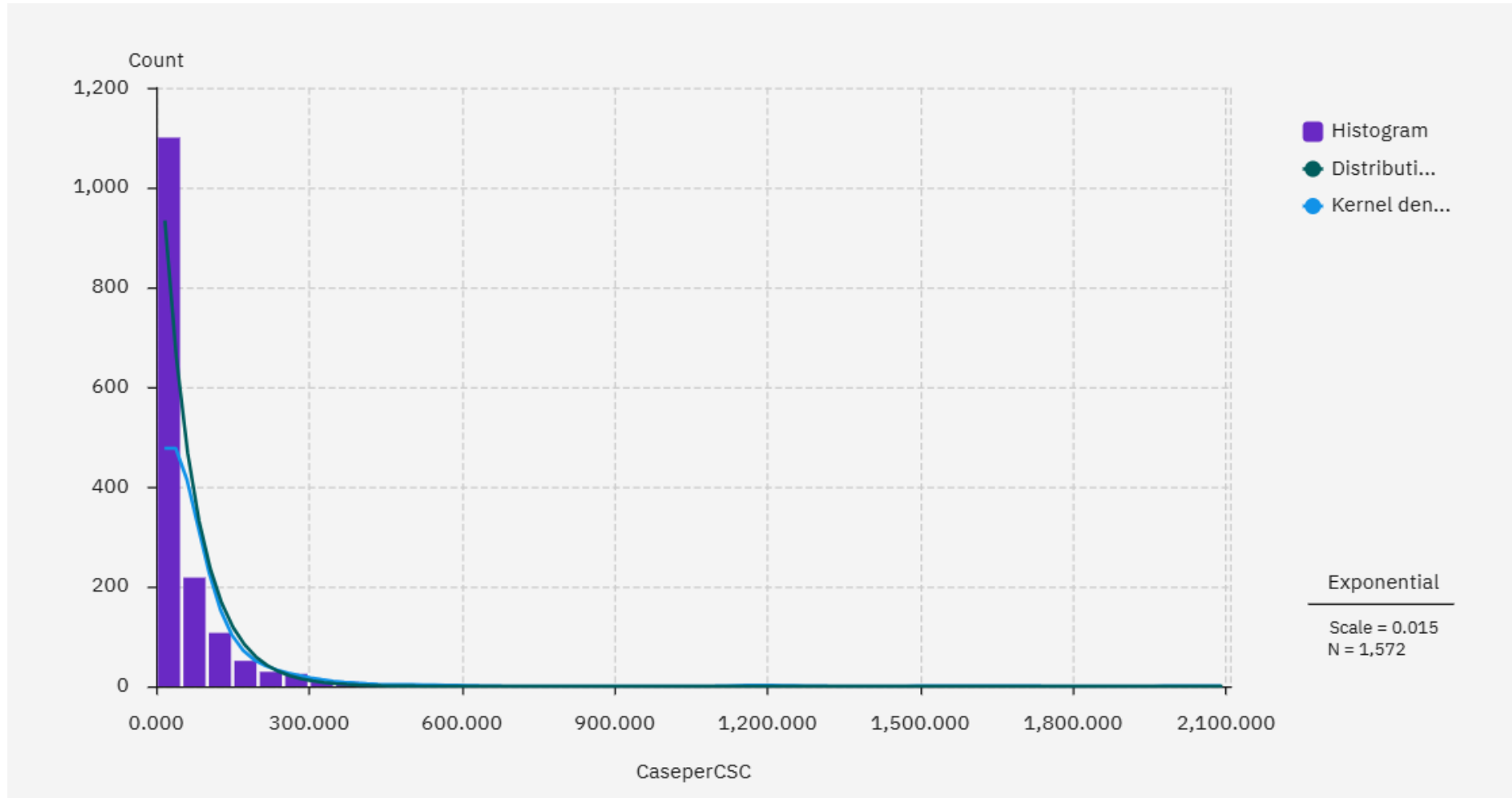
- **Challenges:**

- Incomplete or inconsistent data in some regions.
- Lack of granularity in social or economic context data.

- **Effectiveness:**

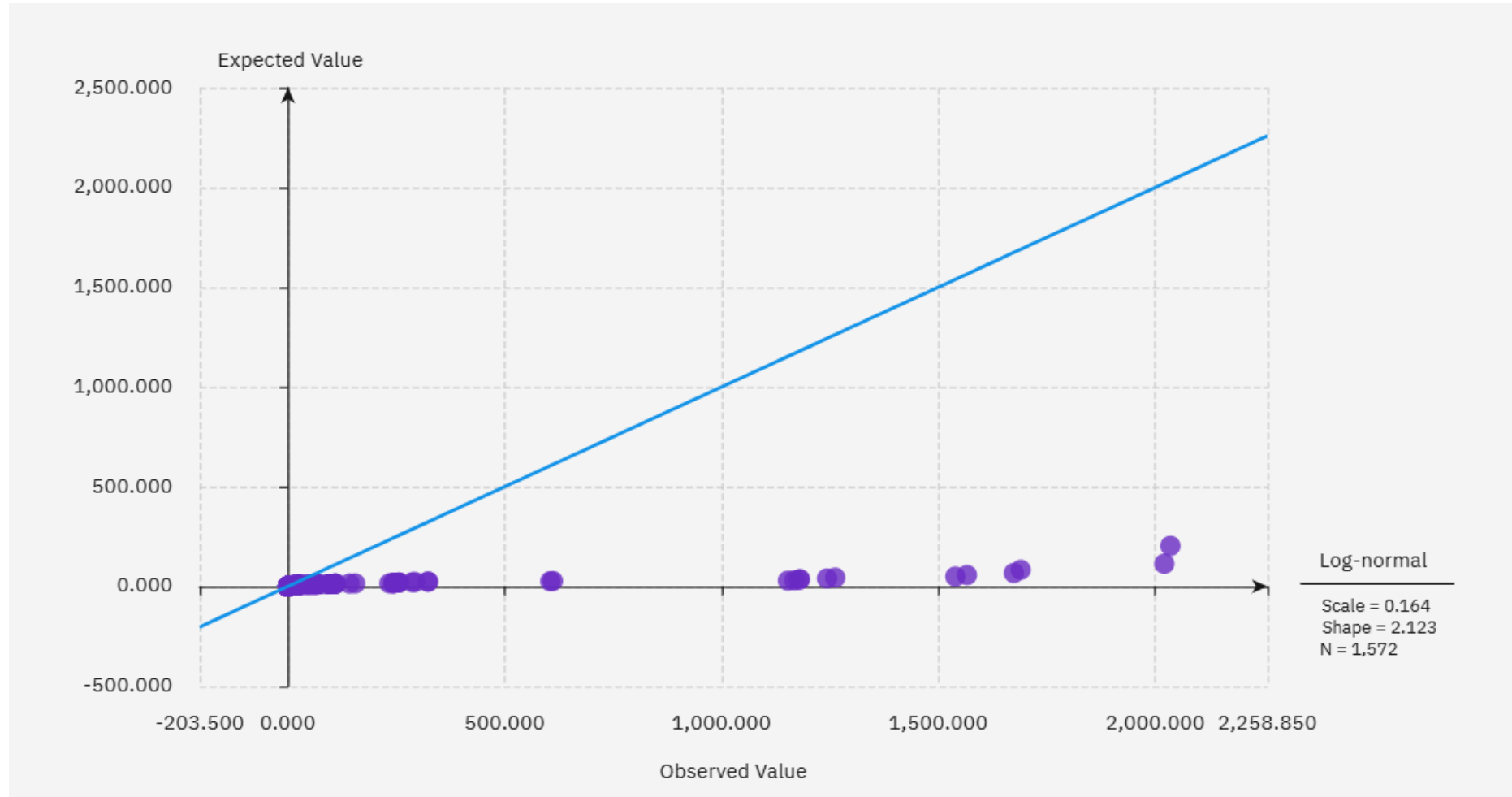
- The analysis effectively uncovers where and among whom legal aid access is lagging.
- These insights can support policymakers and legal outreach programs in targeting resources to the most underserved populations.

CONCLUSION



TOTAL NO. OF CSC = TOTAL / NO. OF CSC

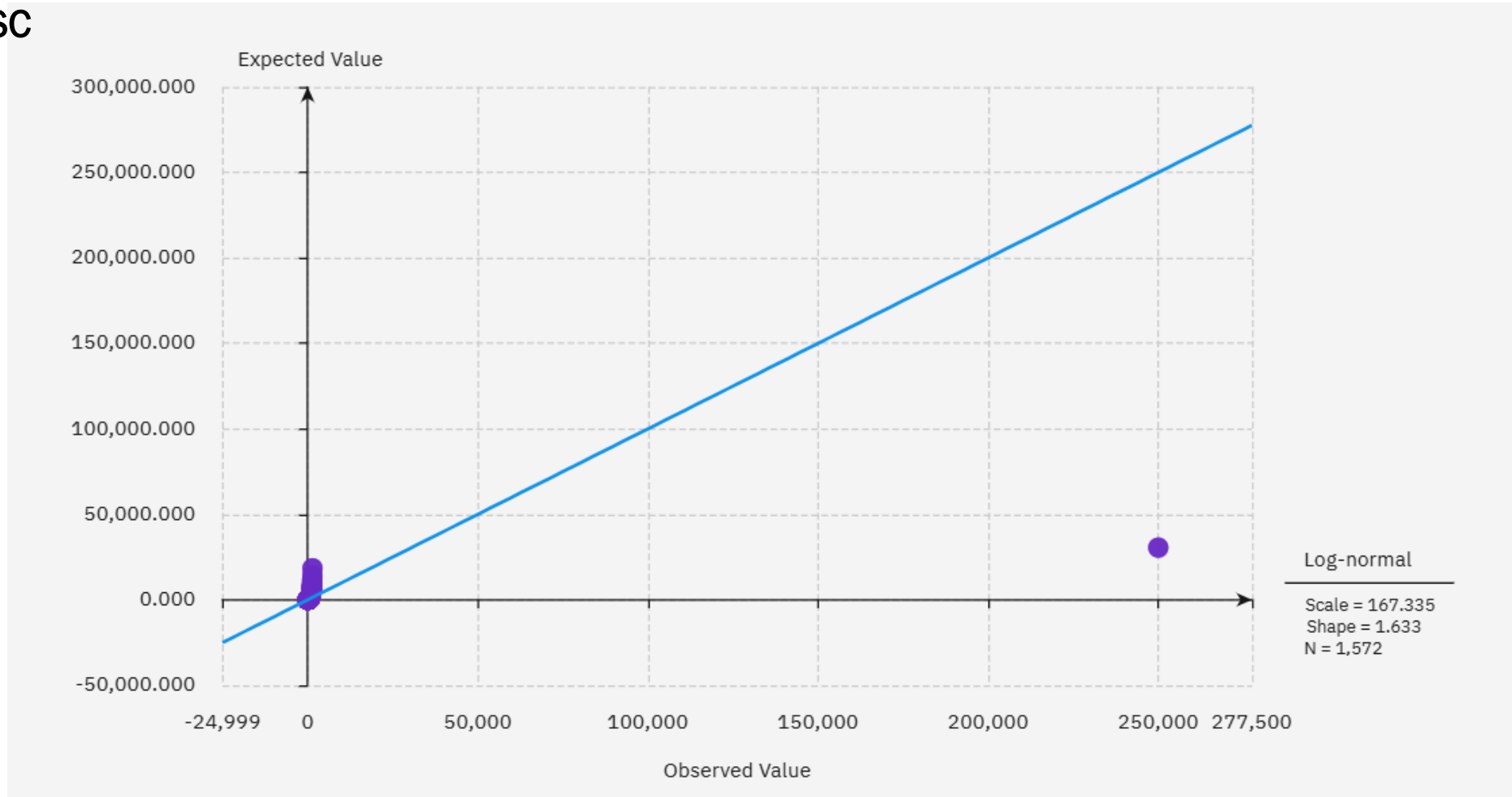
CONCLUSION



CSC=CATEGORY/NO.OF CSC

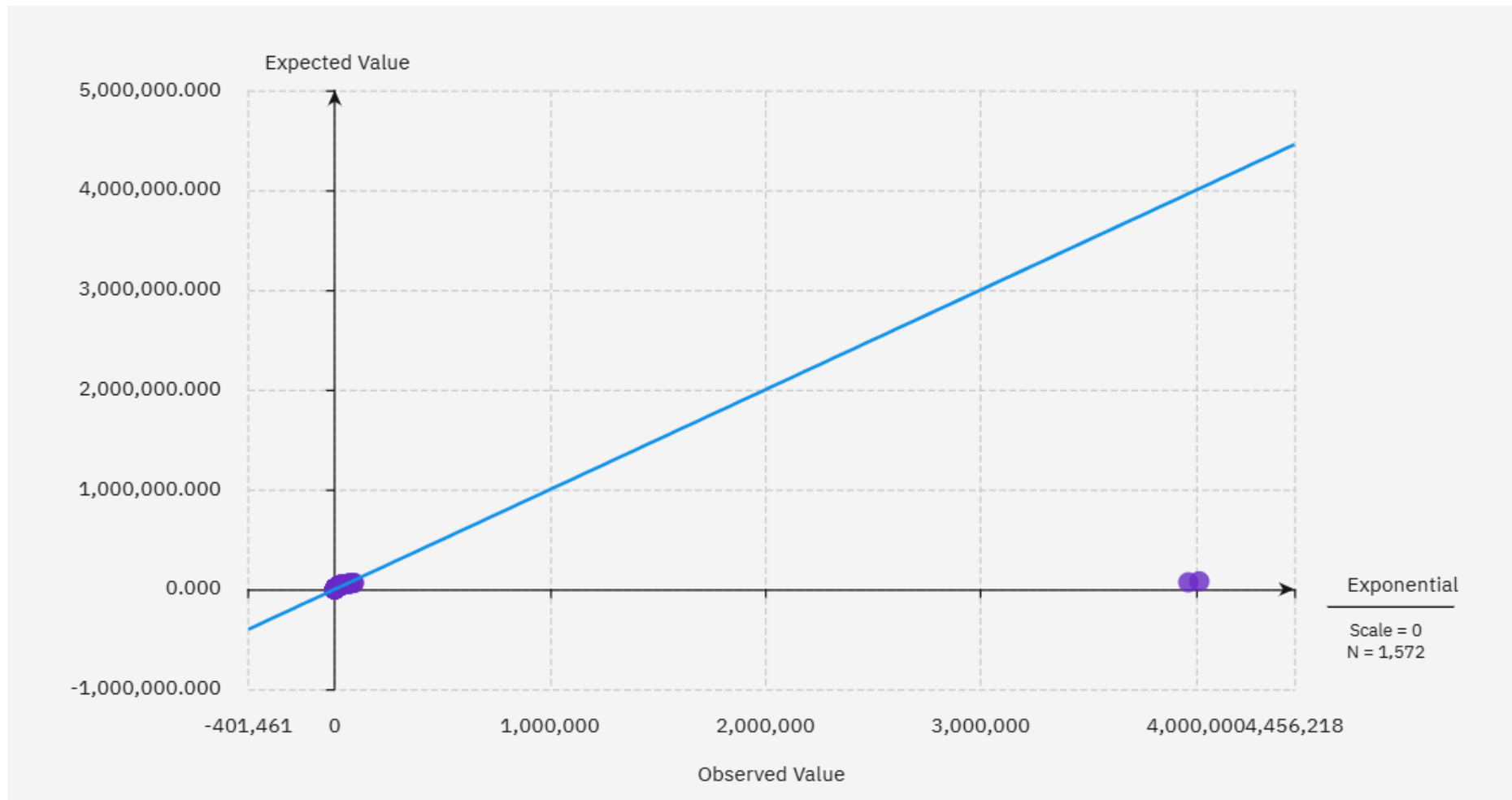
CONCLUSION

NO. OF CSC



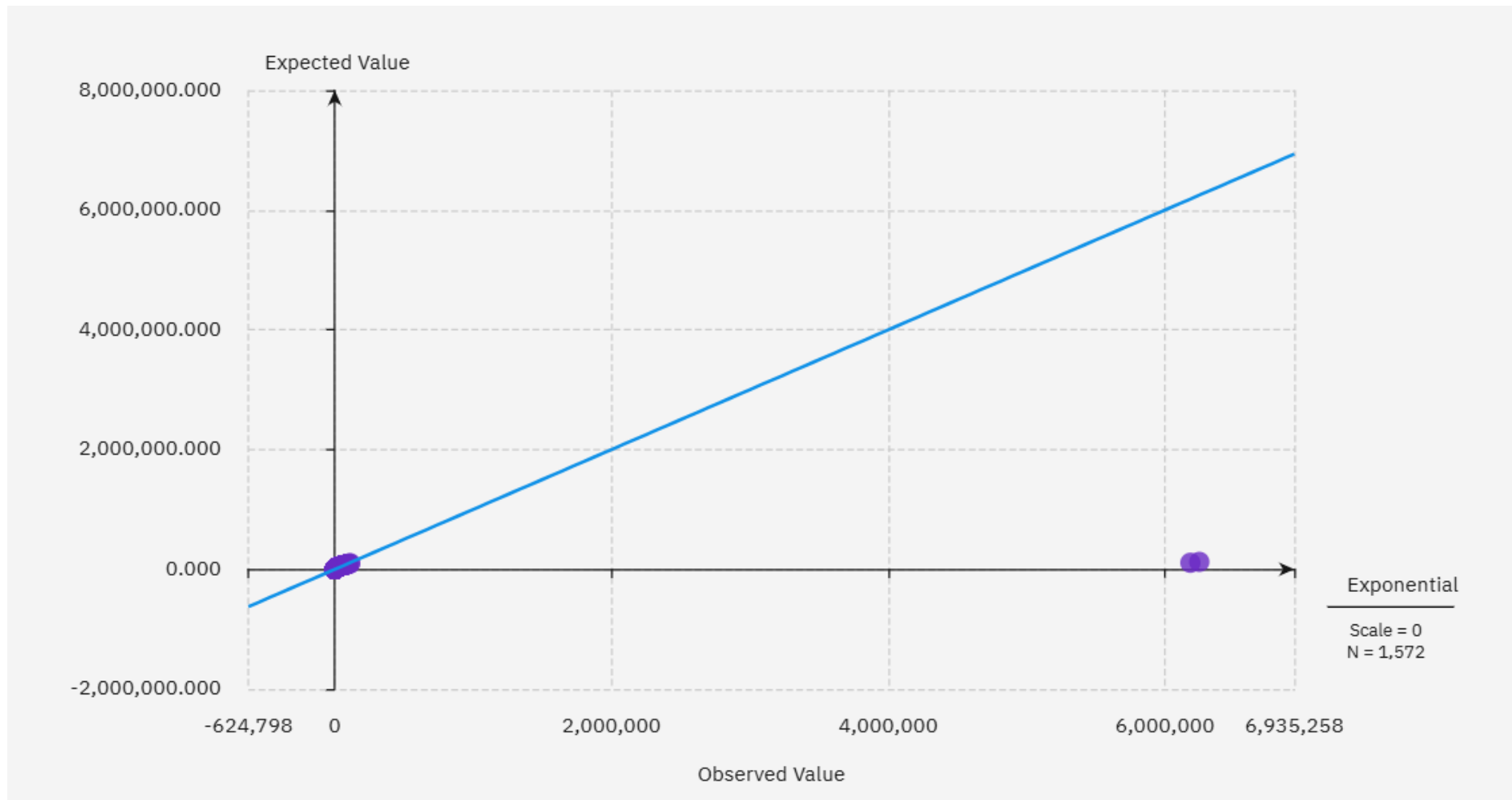
CONCLUSION

FEMALE



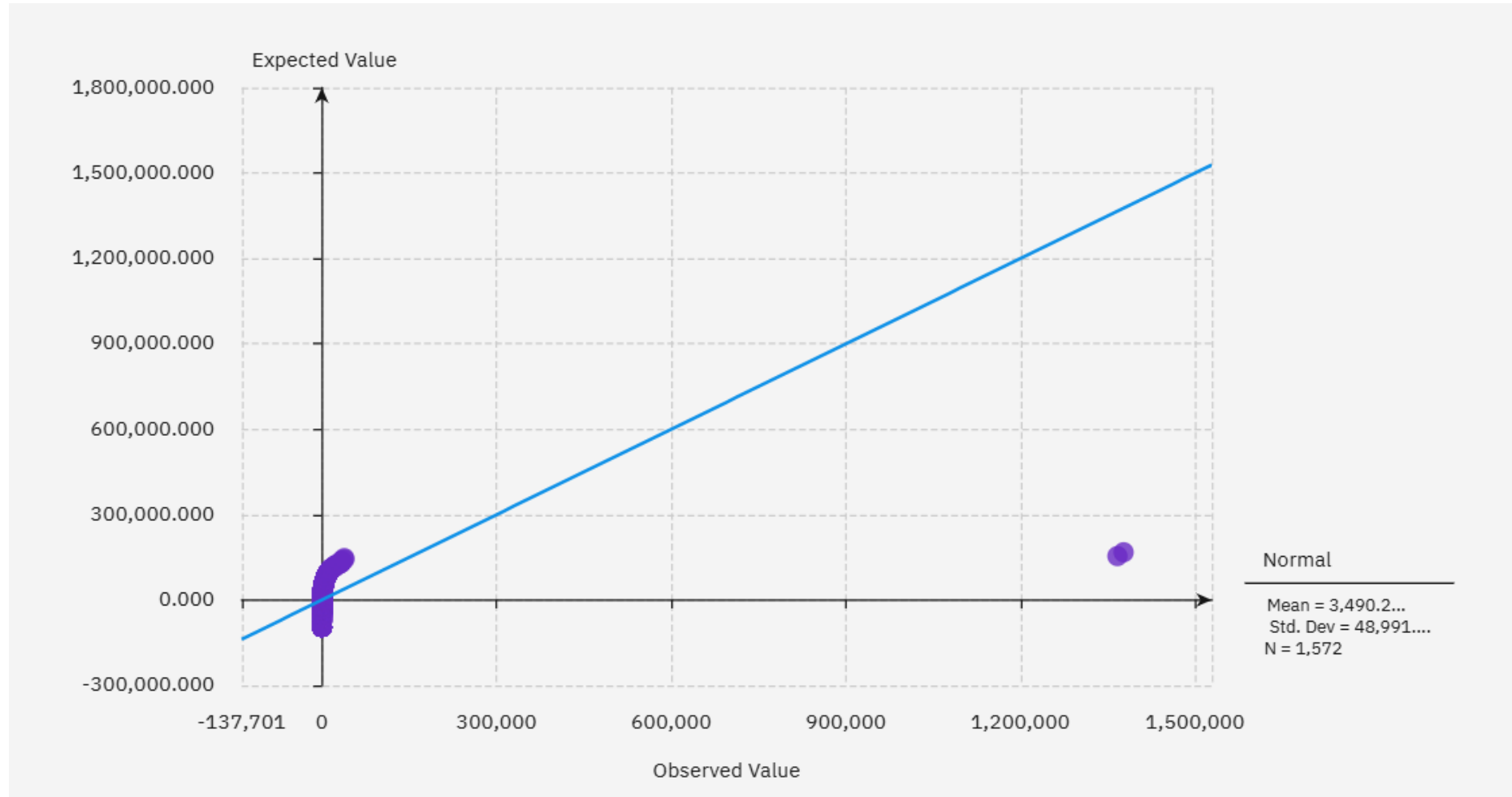
CONCLUSION

MALE



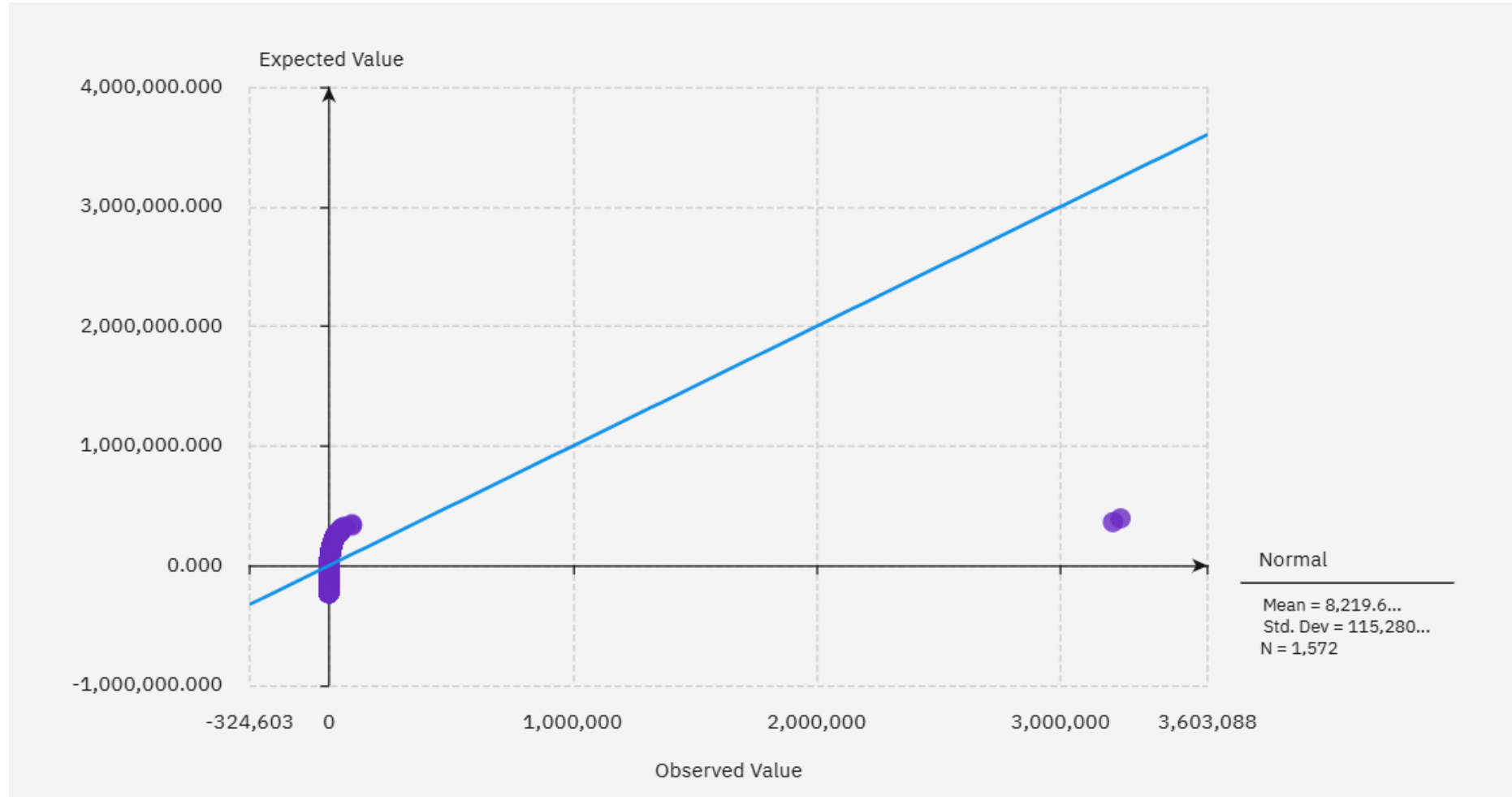
CONCLUSION

GENERAL



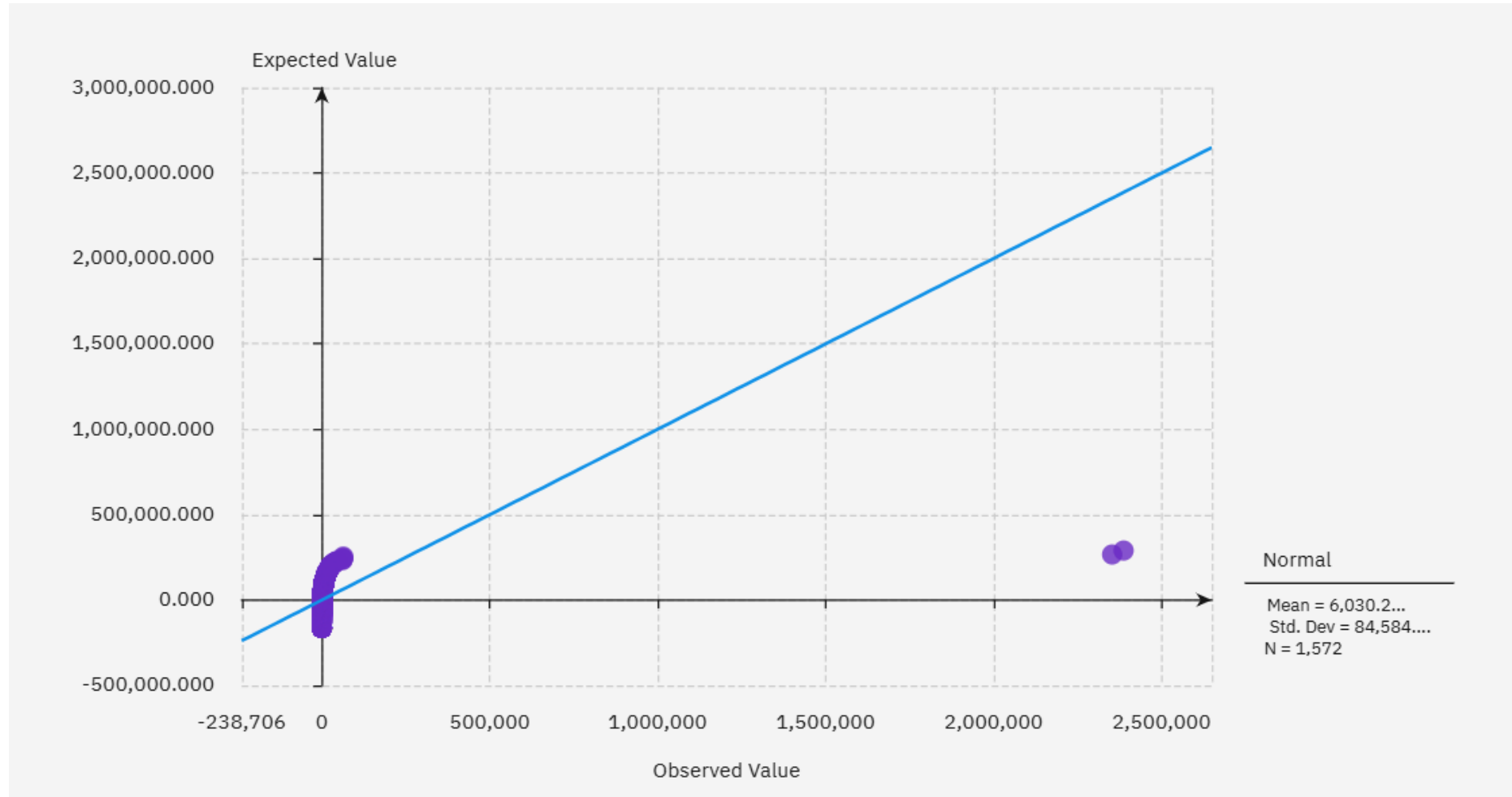
CONCLUSION

SC



CONCLUSION

ST



FUTURE SCOPE

- **Future Scope**
- **Additional Data Sources:** Integrate census data, digital literacy, education levels, and internet availability for deeper insights.
- **Algorithm Optimization:** Implement clustering algorithms or ML models to identify priority outreach zones.
- **Multi-Region Expansion:** Apply the framework to other legal aid schemes and e-governance programs.
- **Emerging Tech Integration:**
 - **Edge Computing:** Enable localized analytics at CSCs.
 - **AI Models:** Deploy predictive models to forecast demand for legal services and optimize resource allocation.
 - **Mobile Integration:** Push outreach content to underrepresented demographics via regional apps.

REFERENCES

- **Relevant Sources and References**
- **Dataset Source :**
 - District-wise Tele-Law Case Registration and Advice Enabled Data (FY 2021-22 to 2024-25)(<https://www.data.gov.in/resource/district-wise-tele-law-case-registration-and-advice-enabled-data-fy-2021-22-2024-25>)
 - Official dataset providing demographic and regional statistics on Tele-Law case registrations across India.
- **IBM Watsonx.ai Documentation :**
 - Watsonx.ai(<https://www.ibm.com/products/watsonx-ai>)
 - For developing AI-driven data analysis pipelines, visualization, and scalable ML workflows.
- **IBM Watson Studio & Cloud Object Storage :**
 - WatsonStudioDocs (<https://dataplatform.cloud.ibm.com/docs/content/wsj/overview/welcome.html>)
 - [Cloud Object Storage Docs](<https://cloud.ibm.com/docs/cloud-object-storage>)
 - Tools used for data preprocessing, storing large demographic CSV files, and integrating dashboards.
- **IBM AutoAI :**
 - AutoAI on Watsonx (<https://www.ibm.com/docs/en/watsonx/saas?topic=services-autoai>)
 - Enables automatic model selection and hyperparameter tuning on Tele-Law datasets for predictive analysis.

REFERENCES

- **Pandas and Seaborn for Python :**
 - Pandas Documentation(<https://pandas.pydata.org/docs/>)
 - Seaborn for Visualizations(<https://seaborn.pydata.org/>)
 - Essential for analyzing caste/gender/district-wise disparities and creating graphical dashboards (bar charts, heatmaps, etc.).
- **Geopandas & Plotly :**
 - GeoPandas Docs (<https://geopandas.org/>)
 - Plotly India Choropleth Maps (<https://plotly.com/python/maps/>)
 - For regional analysis and visualization of Tele-Law cases per state/district.
- **Jupyter Notebooks in IBM Watson Studio :**
 - Using Jupyter in IBM Cloud (<https://www.ibm.com/cloud/blog/how-to-use-jupyter-notebooks-in-ibm-watson-studio>)
 - Interactive development environment for integrating Tele-Law data analysis.
- **Research Articles / Policy Frameworks :**
 - Ministry of Law & Justice – Tele-Law Dashboard](<https://tele-law.in/>)
 - Government’s official initiative for expanding access to legal advice.
 - NITI Aayog and CSC SPV Reports – For rural service distribution strategies and digital inclusion.

IBM CERTIFICATIONS

credly certificate(getting started with AI)

In recognition of the commitment to achieve
professional excellence



Poorva Jain

Has successfully satisfied the requirements for:

Getting Started with Artificial Intelligence



Issued on: Jul 17, 2025

Issued by: IBM SkillsBuild

Verify: <https://www.credly.com/badges/d8e33280-6e5f-4d35-b35e-5f6b3a5b3de3>



IBM CERTIFICATIONS

credly certificate(Journey to Cloud)

In recognition of the commitment to achieve
professional excellence



Poorva Jain

Has successfully satisfied the requirements for:

Journey to Cloud: Envisioning Your Solution



Issued on: Jul 17, 2025
Issued by: IBM SkillsBuild

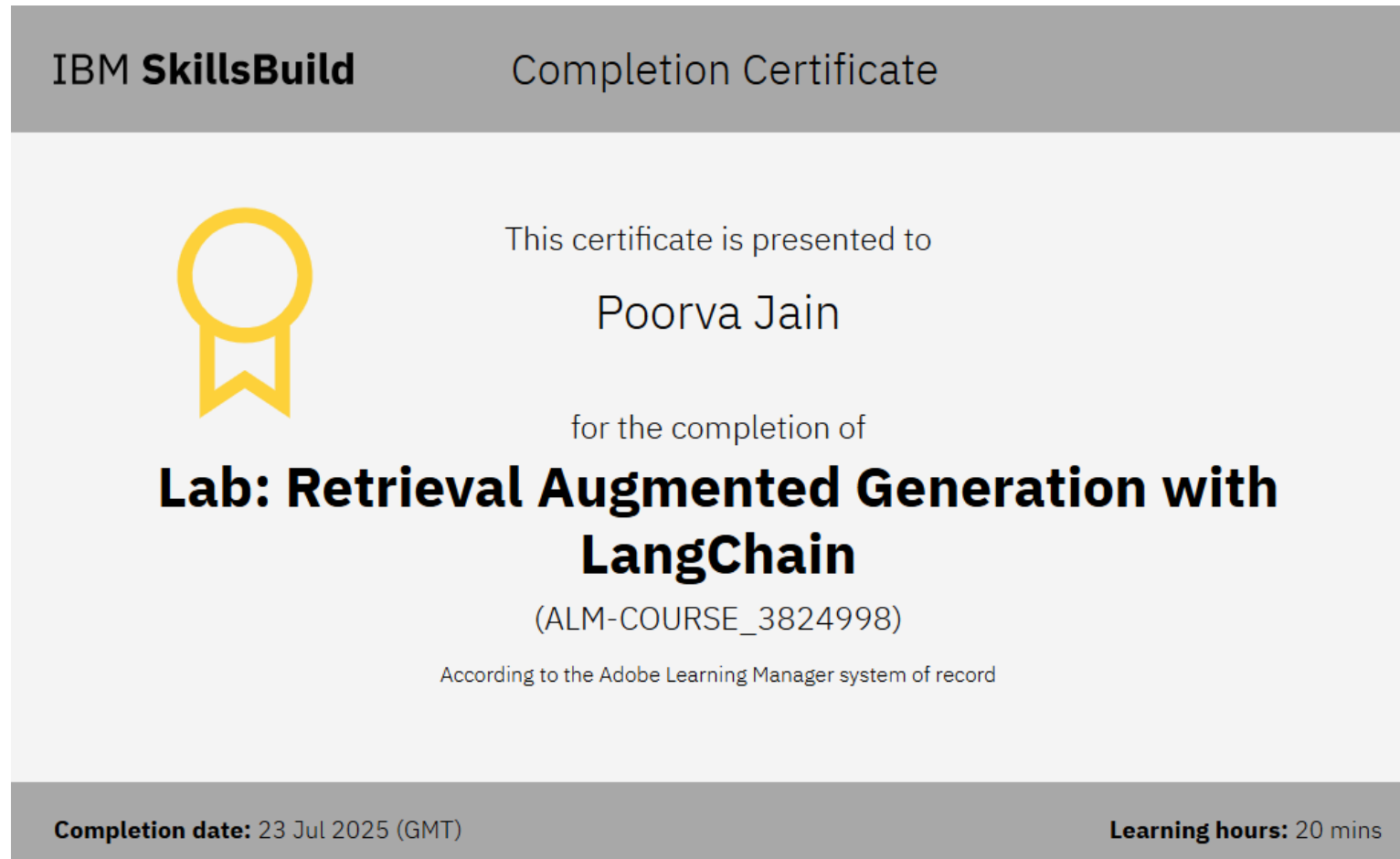
Verify: <https://www.credly.com/badges/2b05a698-0f69-4a6a-91d7-39522c24bf09>



IBM CERTIFICATIONS

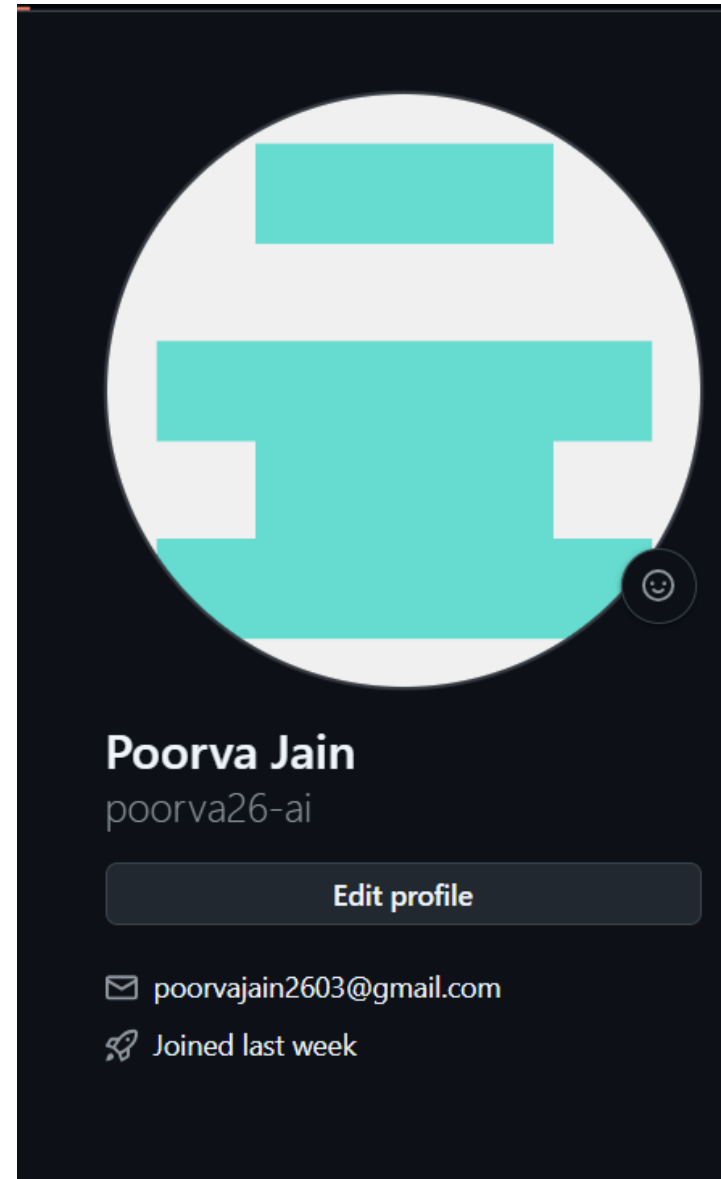
- Screenshot/ credly certificate(RAG Lab)

<https://skills.yourlearning.ibm.com/certificate/share/22c57c6649ewogICJvYmplY3RUeXBliA6ICJBQ1RJVklUWSIsCiAgImxIYXJuZXJDTIVNliA6ICI1MTA4MzU5UkVHliwKICAib2JqZWNOswQilDogIkFMtS1DT1VSU0VfMzgyNDk5OClkQ5a36d0e3aa-10>



GITHUB

- Github : <https://github.com/poorva26-ai>





THANK YOU