Ahuitz Rojas, Ph.D.

ahuitzrojas91@gmail.com

Education

Université Grenoble-Alpes Grenoble, France
Ph.D. Cognitive Science, Specialized in Computational Social Psychology
Institute for Executive Education Puebla, Mexico
M.Sc. Information Technologies Engineering
Universidad del Valle de Mexico San Luis Potosi, Mexico
B.Sc. Psychology

Relevant Experience

Reddit (Akvelon contractor) Remote
Data Scientist/Engineer Mar 2022 - Dec 2024

- Optimized Data Infrastructure: Identified inefficiencies in data retrieval processes and implemented pre-processing views and daily partitions, reducing query times by 90%. This improvement enhanced system performance, enabling faster, more reliable insights for Reddit's consumer-contributor team.

- Enhanced System Scalability: Designed and implemented fault-tolerant, scalable ETL pipelines for Reddit Recap, optimizing data ingestion and transformation. Ensured 99.9% data accuracy and improved system efficiency, allowing backend teams to serve millions of users without performance bottlenecks.

- Increased Product Adoption: Designed and executed A/B tests to assess the impact of new features on business tool usage, post activity, and user interactions for Reddit Pro. These insights shaped product strategy, resulting in higher engagement with analytics tools and positioning Reddit Pro alongside industry competitors.

Used languages and technologies: Python, R, Airflow, BigQuery, GCP, Mode, Hex, Looker, Tableau, Docker, Terraform, Github, Anomalo.

Servicios de Salud de San Luis Potosí (SSSLP)
Data Scientist/Engineer
Jun 2020 - Mar 2022

- Developed Data Pipelines: Built scalable data pipelines using Python, Airflow, Docker, and Terraform to automate data collection. This reduced manual data collection time by 50% and improved system scalability for future data ingestion.

- Developed KPIs and Dashboards: Collaborated with the public health team to define KPIs and design interactive dashboards for timely pandemic updates. Enabled real-time data access, improving decision-making speed and clarity, and reducing reporting time by 90%.

- Created Data Visualizations for COVID-19 Press Conferences: Produced 400+ static visualizations and interactive dashboards using Dash and Shiny for daily press conferences. Automated data processing and visualization reduced workflow time from 12+ hours to full automation, enhancing team efficiency.

Used languages and technologies: Python, R, Airflow, Seaborn, Dash, Shiny, Docker, Terraform, Github.

Université Grenoble-Alpes
Graduate Researcher
Jan 2016 - Dec 2019

- Led 10+ Mixed-Methods Research Studies: Collected data from 1,000+ participants using web scraping, experience sampling methods, experiments, surveys, and interviews.
- Studied Attitude Propagation Across Social Networks: Used web scraping, data mining, and machine learning to analyze social media and discovered the spread of stigmatizing attitudes, particularly in fitness and weight-shaming accounts, highlighting the need for content moderation.
- Communicated Research Insights: Presented insights in 3 empirical research papers and 20+ presentations at scientific conferences across 15+ countries, in 3 languages.

Used languages and technologies: R, Python, Scikit-Learn, Tidyverse, Github.

SKILLS & STRENGTHS

Tools:
Python, R, Apache Spark, Airflow, SQL(PostgreSQL, MySQL, BigQuery), Data visualization (Tableau, Mode, Matplotlib, Seaborn, ggplot2), Docker, Infrastructure as code (Terraform), Cloud Computing (GCP), Git and github, Linux/Bash Scripting.

Skills:
Data analysis, machine learning, hardening pipelines through monitoring and optimization, KPI design and anomaly detection, statistical modeling, A/B testing, causal inference, experimental design and data modeling.

Languages:
Spanish - Native, **English** (C2/native proficiency), **French** (C1/advanced)

Relevant Projects

- Structural Equation Modeling (SEM): Applied SEM as a causal inference tool to assess the impact of discrimination on health outcomes. Cross-cultural invariance was confirmed between French and Mexican samples with 68% certainty. [Link]
- Network Analysis and NLP on Twitter: Leveraged Twitter's API, Network Analysis, and Natural Language Processing to examine the overlap between health promotion and weight-based discrimination. The combined network (degree 12.91, path length 4.31) showed improved connectivity and a reduced diameter, suggesting fitness-talk users act as bridges between fragmented weight-talk communities. [Link]
- Web-Scraping and NLP: Utilized web-scraping, Network Analysis, and NLP to uncover a network of individuals sharing images of sexual abuse in Mexico. The network reaches over 130,000 people, with consumers of abusive content being the primary drivers of its spread. [Link]