Pragati Prashant Ingole

24pragatipingole@csu.fullerton.edu | (657)-799-8731

https://www.linkedin.com/in/pragatiingole/ | Github | Portfolio

STHRENGTHS AND EXPERTISE

Languages: Python (NumPy, Pandas, Matplotlib, Seaborn), R (ggplot2, Shiny), SQL, VBA, Shell Scripting C, C++, HTML, CSS
Software/Tools: Tableau, PowerBI , Excel (VLOOKUP, Pivot tables, VBA, chart functions)**,** JIRA, Alteryx, Adobe Analytics.
Database Systems: PostgreSQL, PG Admin**,** SAS, SPSS, MongoDB, MySQL, RDBMS
Version Control/ Cloud: Git, GitHub, Amazon Web Services
Statistical Techniques: Hypothesis Testing, Regression, ANOVA, A/B Testing.
Skills: Data Analysis, Database Systems, Data Modeling, ETL Procedures, Data Manipulation, Data Cleaning, Data Transformation, Reporting and Visualization, Requirements Gathering, Marketing Analytics, SDLC, Relational Databases, Strong communication and Presentation skills.
Certification: Google Data Analytics Professional Certificate,

PROFESSIONAL EXPERIENCE

Graduate Data Analyst Assistant of IT Dept. Data Center **Sep 2023 – Present**
California State University, Fullerton, CA

- Developed Python and PySpark MLlib (time series forecasting, K- means) algorithms using ARIMA and Prophet to analyze patterns and anomalies in server load performance trends, boosting key metrics by 20% through enhanced data analysis frameworks.
- Designed and executed robust data extraction and processing pipelines and developed ETL workflow using Structure Ware, Rubix, culminating in a dynamic dashboard with Tableau that enhanced inventory forecasting accuracy to 99%. Utilized APIs for real-time university student portal data updates and ensured data integrity across multiple sources.
- Enhanced database query efficiency by 25% by optimizing and batching SQL Alchemy operations (indexing, partitioning, and query rewriting) using Python scripts. Used multi-threading and connection pooling to manage concurrent database connections, significantly reducing server load and processing time.

Graduate Research Assistant **Jul 2023 - Sep 2023**
California State University, Fullerton, CA

- Advanced a high-accuracy data collection system using CATI software, designing data entry interfaces and validation checks, processing over 10,000 student records with a 99.5% accuracy rate and compliance with university standards.
- Executed data preprocessing workflows using Python (pandas and NumPy, handling missing values, scaling, normalization, and one-hot encoding) creating automated workflows, enhancing data quality by 20%, enabling effective exploratory data analysis and pattern.
- Piloted quantitative research through statistical analyses using R and SPSS, coordinating with departments to gather and integrate data, presenting insights that influenced educational strategies, boosting student retention by 15%.

Associate Consultant **Apr 2022 - Aug 2022**
Bristlecone Pvt Ltd, Pune, India

- Worked with client Amazon to design and implement a cloud-based billing system, improving financial data handling efficiency by 10%, using AWS Lambda, RDS, S3, IAM for security, and Step Functions for workflow automation.
- Optimized SQL data migration for 750+ retail transactions** by implementing indexing and partitioning strategies, using batch processing and bulk inserts, applying normalization techniques, and optimizing query execution for data integrity and strategic decision-making.
- Integrated multiple data sources into Power BI, creating dynamic dashboards, writing complex DAX queries, and automating data extraction with Shell scripts; implemented Tableau and Excel solutions, increasing data-driven decision-making practices by 30%.

Idle Solutions Ltd. Pune, India **Aug 2021 – Mar 2022**

- Oversaw end-to-end delivery of a product line's data load, from Data Modeling to pipeline using Power BI and Python transformation to integrate data sources into a cohesive pipeline. Added Change Data capture (CDC) via Quilkview to capture daily changes, ensuring timely and accurate data delivery.
- Collaborated with cross-functional teams, delivering a $500,000 project by translating business needs into automated ETL pipelines using Informatica Power Center streamlining data flow, integration processes. Reduced loan processing time from 40 days to one week.
- Utilized Excel and SQL, deploying PL/SQL codes, stored procedures, views, functions, and window functions to ensure data integrity, automate ETL processes, and optimize query performance for advanced data analysis.
- Loaded data from Enterprise Data Warehouse into Salesforce via Informatica, ensuring seamless integration and data accuracy. Managed migration maintained data quality, and validated data in the nCino application for reliable data entry.
- Automated Informatica jobs using UNIX SHELL SCRIPT which led to reduction in manual resources by 40%.
- Streamlined KPI tracking by creating automated PowerBI dashboards using DAX and Python for data manipulation, which included fetching data from APIs and scheduling dashboard refreshes for accuracy and currency.

ACADEMIC PROJECTS

Diabetes Health Indicators Analytics [Tableu, Python] **May 2024**

- Analyzed 253,680 health records to identify an inverse correlation between physical activity and diabetes prevalence, demonstrating a 15% reduction in diabetes risk with increased physical activity.
- Developed logistic regression and decision tree models with 95% accuracy to predict diabetes incidence, identifying high blood pressure, high cholesterol, and BMI as key risk factors.

Stocks Portfolio Optimization [SQL, PostgreSQL, R, Excel] **Nov 2023**

- Built a stock portfolio against benchmark data SP500TR and devised a data analysis methodology to identify patterns in student dropout rates, discovering a critical trend where 90% of dropouts had attended intervention seminars.
- Developed and interpreted multi-dimensional data tables to forecast dropout probabilities, providing actionable insights for educational improvement.

Business Intelligence System for Enhanced Product Profitability and Market Analysis [ETL, Snowflake] **July 2023**

- Conceptualized tables and columns are required to make the data warehouse ready to answer any data analysis questions. Designed a data warehouse with **Snowflake** one fact table and four-dimensional tables.
- Developed **ELT design** document with two ELT pipelines to ingest the incoming data for Log tables: Day 0 load to load all the history data into the data lake and DayN load to load only incremental data into the **data lake**.

EDUCATION

Master of Science, Management Information Systems (Business Analytics) **Aug 2022 - May 2024**
California State University Fullerton, California, USA **GPA-3.57**
Bachelor of Engineering (Electronics and Telecommunications) **Aug 2016 – May 2021**
University of Pune, India