Sandeep Kumar

Sk175@iu.edu | +1 812 803 5058 | LinkedIn

EDUCATION

Indiana University - *Master of Science in Information Science (GPA 3.7/4)* — *Bloomington, IN | Dec 2024*
S.G.S.I.T.S - *Bachelor of Engineering in Electrical (GPA 7.3/10)* — *India | June 2019*
Certifications - *PowerBI, Power Automate, SCRUM, PMP*
Relevant coursework - *Statistics, Data Fluency, Applied ML, Data visualization, Social media mining, Data Fluency, DB design.*

PROFESSIONAL EXPERIENCE

Spirent Communication - Santa Clara, CA — *June 2024-Aug 2024*
Business Process Automation Intern

- Implemented **Approval Automation flow** using **Power Automate** to streamline approval processes across various business units.
- Conducted **Exploratory Data Analysis (EDA)** using **python** to identify bottlenecks and inefficiencies, reducing cycle time by 25%.
- Developed **Tableau dashboards** to visualize KPIs and create heat maps, leading to a 20% reduction in approval processing time.
- Implemented data cleaning procedures using **Python** and **SQL**, improving data accuracy and reliability by 29% for approval workflows.

Indiana University - Bloomington, IN — *July 2023-Dec 2024*
Teaching Assistant (Statistics)

- Developed and optimized 20+ **R** and **Python** guides for advanced **data analytics** and **statistical methods**, supporting academic research publications by automating data cleaning, transformation, and visualization processes.
- Utilized **SQL** for querying large datasets and **Python (Pandas, NumPy)** and **R** for implementing complex statistical models and data manipulation techniques, contributing to cancer research by enhancing the extraction and analysis of high-volume datasets.
- **Streamlined data cleaning** by automated scripts in **Python** for large datasets, removing inconsistencies and handling missing values.
- Developed compelling data visualizations using **Python (Matplotlib, Seaborn)** and **R (ggplot2)** to present findings in research reports.

Decision Point Analytics - Delhi, India — *Jul 2022-Dec 2022*
Program Manager

- Integrated diverse data sources into data warehouses, built **ETL pipelines** in **Python** using techniques like data cleaning, normalization, and incremental loading, and used **SQL databases** for data modeling and processing to streamline workflow.
- Spearheaded the development of a Resource Management Tool in **Power BI**, improving transparency & project efficiency by 11%.
- Optimized 60+ **SQL** queries using **query optimization** and **indexing**, applying **normalization** principles, reduction in execution time.
- Developed **Power BI** dashboards and self-service tools, leveraging **DAX** for advanced calculations and **SQL** for data integration, allowing senior management to monitor trends in performance metrics leveraging and **agile** methodologies within the **SDLC** framework.
- Developed **database schemas** using **Microsoft Visio**, applying **entity-relationship modeling (ERD)** and **normalization** techniques.

VOLVO — *Jul 2019-Jul 2022*
Project Manager / Data Analyst

- Managed project estimates and forecasts using **Excel** for modeling and scenario planning, and **Python (Pandas, NumPy)** for automated data validation, exceeding performance expectations by 5% and reducing data validation time by 17%, earning the 'Innovation Award'.
- Automated status of order reports for the production planning team using **Excel VBA scripts** and **macros** to merge, format, rename, separate, and sort columns, reducing daily report creation time by 30 minutes each morning.
- Examined labor data using **seaborn** radar, scatter plots, uncovering workload imbalances, improving task distribution by 21%.
- Utilized **Power BI** and **SQL** to support **business intelligence** and reporting across 10+ manufacturing units, employing **ETL processes** and **data visualization** techniques while leading cross-functional collaboration to track project progress and resolve issues.
- Crafted 20+ **KPIs** per assembly line using **DAX, Power BI**, and **Python** to track productivity targets, enabling informed decision-making on employee scheduling to ensure coverage of required productivity hours and enhance overall efficiency.

PROJECTS

- **Statistical Modeling**: Data-driven model in R to optimize FIFA team selection, employing PCA, clustering, and predictive modeling.
- **Social Media mining**: Utilized Python (PRAW, Pandas) and Reddit API to scrape & analyze 1,000+ posts, using sentiment analysis and topic modeling to uncover 38 distinct themes. Applied natural language processing techniques to reveal thematic trends.
- **Churn Prediction & Revenue Forecasting**: Built classification models (Logistic Regression, Decision Trees, Bootstrap Forest) to predict churn with 94% accuracy, applied Lasso Regression for revenue prediction, enhancing performance by feature selection.
- **Database Design**: Designed a 3NF-normalized MySQL database with optimized data types and indexed queries, enabling millisecond access to job listings. Integrated CRUD operations with Flask, enforcing referential integrity and reducing data latency by 28%
- **Guaranteed Income for Artists Program Analysis:** Utilized Python (Pandas, NumPy, Matplotlib, Seaborn) for data analysis and Adobe Illustrator for visualizations, winning first prize in a data analysis competition for unique insights and artistic visualization.

SKILLS

- **Programming Languages**: SQL, Python (Pandas, NumPy, Scikit-learn, Matplotlib), R (dplyr, ggplot, tidyr), SAS, DAX
- **Data Analysis Tools**: Tableau, Power BI, Looker, Excel (Pivot Tables, VLOOKUP), VBA, Looker, Google Analytics
- **Databases and Web Technologies**: MySQL, NoSQL, PostgreSQL, Neo4j, BigQuery, Snowflake, HTML5, JavaScript
- **Data Management**: Data Wrangling, ETL processes, Data Warehousing, Database Management
- **Statistical Analysis**: Hypothesis Testing, Regression Analysis, Survival Analysis, Network Analysis, GLM,
- **Machine Learning**: Predictive Modeling, Classification, Clustering, Data Mining
- **Cloud Platforms**: AWS (S3, EC2, Lambda, SageMaker), Google Cloud Platform (GCP), Microsoft Azure
- **Misc**: Jupyter Notebook, Data Structure, NLP, Hadoop, Spark