**Sagar Ippili**

**Chicago, IL – 60616 ● (312) 539-2388 ●** [**sippili@hawk.iit.edu**](mailto:sippili@hawk.iit.edu) **●** [**linkedin.com/in/sagarippili**](https://www.linkedin.com/in/sagarippili) **●** [**github.com/sagarippili**](https://github.com/sagarippili)

**Objective**

Looking for an opportunity as a Data Analyst/Scientist to better analyze, design and implement data science concepts that make a difference and thus, contributes to the company’s growth

**Education**

|  |  |
| --- | --- |
| * Illinois Institute of Technology, Chicago, United States | May 2020 |
| **Master’s: Information Technology | Data Science & Management** | **GPA: 3.88** |
| * Gandhi Institute of Technology and Management, India | Apr 2014 |
| **Bachelor’s: Electronics & Instrumentation Engineering** | **GPA: 3.8** |

**Skills**

* **Languages/Scripts:** Bash Shell Scripting, Perl, Python, R, SQL
* **Data Science | Analytics:** Polynomial Regression, Classification, Time Series, Clustering, Predictive Modeling, Pandas, NumPy, SciPy, SciKit-Learn
* **Data Visualization:** Tableau, Power BI, Folium, Seaborn, Matplotlib, Forecasting
* **Big Data Technologies:** HDFS, Hadoop/MapReduce, Spark, Hive, Pig
* **Databases:** Oracle 12c, MySQL, MS Access, PL/SQL, Excel
* **Tools, IDE, Servers:** Jupyter Notebook, Linux/UNIX, Visual Code, Visio, SharePoint, Pentaho, ETL, GIT, Xen Virtualization, High Availability, Disaster Recovery, Jenkins
* **Cloud:** AWS EC2, AWS S3, AWS Redshift, AWS VPC, AWS ELB, AWS AMI, IBM Cloud, IBM Watson Studio

**Certifications**

|  |
| --- |
| * **Tableau Desktop Associate by Udemy** |
| Credential ID: UC-50616b00-07ca-495e-9eb0-5cdb51afbd4d |
| * **IBM Certified Data Analysis with Python** |
| Credential ID: Coursera RVZZF52Z56MZ |
| * **IBM Certified Data Visualization with Python** |
| Credential ID: Coursera 2DJKEA44DVWW |

**Professional Experience**

|  |  |
| --- | --- |
| **Data Analyst Intern** | May 2019 – Apr 2020 |
| Chicago Transit Authority, Chicago, United States | |
| * Cleaned the 100,000 (approx.) records of unstructured legacy data using Python’s data wrangling techniques with Pandas, Numpy. * Analyzed the Records Center’s data to generate ad-hoc reports per requirements, using SQL, Amazon Redshift, Excel, and Access. * Improved the overall cross inventory search by 85% using Power BI dashboarding techniques and modeled a search application using the PowerBI reports and slicers/filters. * Published and scheduled automated refreshes using the PowerBI Gateway connecting Desktop to the SharePoint website. * Programmed SharePoint workflows to automate processes like Request and Disposal of CTA’s records with RESTful API services. | |
| **System Engineer** | Nov 2014 – Jul 2018 |
| Tata Consultancy Services Limited, India | |
| * Coded in Perl, Shell, and Python, and documented artifacts including designs, unit test plans, thus contributing to hands-on experience in every stage of an SDLC in an Agile driven project. * Optimized a parameter handling tool using stored procedures in PL SQL for improving the efficiency of the tool by 40%. * Installed, upgraded and deployed Linux SuSE physical and virtual machines over live network hosting 1000+ network elements including SAN and LUNs configuration for High Availability clustering setup using Linux command line. * Developed Shell scripts that are programmed to automate the legacy procedures to back up and restore Linux physical machines and virtual machines configured in a High Availability cluster. Thus, it reduced the downtime of live servers by 60%. * Developed an application using AWS core components like Amazon EC2, S3, ELB, AutoScaling, IAM. Worked with AWS CLI and Cloud Formation Templates and scripts in python using the Boto3 framework to interact with web services. * Re-engineered the installation of High Availability solutions on Linux machines using Python to reduce the overall time to set up by 40%. | |

**Academic Projects**

|  |  |
| --- | --- |
| * **COVID-19 Cases: Tableau, GIT** | **Mar 2020** |
| Built a dashboard visualizing the COVID-19 cases to provide meaningful insights on the daily cases using the dataset from Tableau COVID-19 Data Resources hub ([tableau.com/profile/sagarippili](https://public.tableau.com/profile/sagarippili)). | |
| * **NYC Parking Violations Prediction: PySpark, Pig, Hive** | **Aug 2019** |
| Performed data analysis on the NYC Parking tickets data from years 2016 – 2019 on Kaggle using different BigData tools and built a recommender system to predict the street where most of the violations occurred. | |
| * **Black Friday Sales Prediction: R, Classification, Tableau** | **Aug 2018** |
| Built different classification models like KNN, Naïve Bayes and logistic regression to find a better model among the three to predict the customer purchase pattern with the highest accuracy using MAE and AIC metrics and Hold out evaluation. | |
| * **Time Series Analysis on Employment Rate: R, Power BI** | **Dec 2018** |
| Built ARIMA, ARMA time-series models to analyze the difference in employment rate based on gender over time and forecast the future rates, performed hypothesis testing to identify gender bias while hiring an employee. | |