|  |
| --- |
| **Sagar Ippili**  **Armonk, NY - 10504 ● (312) 539-2388 ● 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 and Management | Data Science & Management**   * Coursework: Data Analytics, Data Mining, Data Warehousing, Big Data Technologies, Advanced Database Management, Python Programming | | **GPA: 3.9/4.0** |
| Gandhi Institute of Technology and Management (GITAM), Visakhapatnam, India | | Apr 2014 |
| **Bachelor’s: Electronics and Instrumentation Engineering** | | **GPA: 3.8/4.0** |
|  | |  |
| **Skills** | | |
| **Languages/Scripts** | : Python, Unix Shell Scripting, Perl, SQL | |
| **Data Science | Analytics** | : Polynomial Regression, Classification, Clustering, Pandas, NumPy, SciKit-Learn, BeautifulSoup, JSON | |
| **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 Studio Code, SharePoint, Pentaho, ETL, GIT | |
| **Cloud** | : AWS EC2, AWS S3, AWS Redshift, AWS VPC, AWS ELB, AWS AMI, IBM Cloud, IBM Watson, IBM DB2 | |
|  | | |
| **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. * Modeled a search application on Power BI that extracts and mines data from SharePoint site to create a dashboard improving the overall cross inventory serach by 85% using slicers and 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, Hyderabad, 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%. | | |
|  | | |
| **Projects** | | |
| * **Apartment Rental Prices Analysis of Chicago: Python, Clustering, GeoSpatial, Webscraping** | | **Apr 2020** |
| Analyzed apartment rental prices in City of Chicago on the basis of neighborhood segmentation using K-Means Clustering, and rental listings scraped from Zillow.com using BeautifulSoup library in JSON format utilitzing IBM Watson Studio and IBM DB2 cloud resources. Visualized maps using Folium, Choropleth and Nominatim. Link to the Jupyter Notebook: [**github.com/sagarippili/ChicagoRentalPricesAnalysis**](https://github.com/sagarippili/Coursera_Capstone/blob/master/Battle_Of_The_Neighborhoods_Capstone_Project/Apartment%20Rental%20Price%20%26%20Venues%20Data%20Analysis%20of%20Chicago.ipynb). | | |
| * **Loan Default Prediction | KNN, Decision Trees, Logistic Regression Classifiers, SciKit Learn, Pandas** | | **Apr 2020** |
| Built different classification models like KNN, Logistic Regression, Decision Trees and SVM to predict loan defaulters and found the best classifier using evaluation metrics like F1-Score, Jaccard Score, Log loss and confusion Matrix. Link to the Jupyter Notebook: [**cloud.ibm.com/analytics/notebooks/LoanDefaulterClassification**](https://dataplatform.cloud.ibm.com/analytics/notebooks/v2/65149db6-87e1-48c1-8931-50184bfc46f9/view?access_token=a3bd8f2046fdaf5625dd5f0263459a14a6a0a85027425c9234d8c567f41331b4). | | |
| * **COVID-19 Cases Dashboard | Tableau, Git** | | **Mar 2020** |
| Built a dashboard visualizing the COVID-19 pandemic’s analysis to provide meaningful insights on the daily cases using the dataset made available to the public by Johns Hopkins University and Tableau COVID-19 Data Resources hub. Link to dashboard : [**tableau.com/profile/sagarippili/COVID-19CasesDashboard**](https://public.tableau.com/profile/sagarippili#!/vizhome/COVID-19CasesDashboard/COVID-19CasesDashboard). | | |
| * **NYC Parking Violations Prediction | Pig, Hive, Spark** | | **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. Link to project code and documents: [**github.com/sagarippili/AvoidingNYCParkingViolations**](https://github.com/sagarippili/AvoidingNYCParkingViolations). | | |
| * **Black Friday Sales Prediction | R, Logistic Regression, Naïve-Bayes, 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. Link to project code and documents: [**github.com/sagarippili/BlackFridaySalesPrediction**](https://github.com/sagarippili/BlackFridaySalesPrediction). | | |
| * **Employment Rate analysis using Time Series | R, Time Series Analysis, 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. Achieved 40% better results using ARIMA model over the ARMA model. | | |
|  | | |
| **Certifications** | | |
| **Tableau Certified Desktop Specialist** | | **May 2020** |
| * Credential ID: 1029416 ([**youracclaim.com/TableauDesktopSpecialist**](https://www.youracclaim.com/badges/adf03ac0-a41c-4695-8eeb-9c4efdefe2eb)) | | |
| **IBM Data Science Professional Certificate** | | **Apr 2020** |
| * Credential ID: Coursera 5ZZMJQYTKS7R ([**youracclaim.com/IBMDataScienceProfCertificate**](https://www.youracclaim.com/badges/b2c3724d-1efd-4c4b-9113-f9d4431461f7)) | | |
| **Tableau Analyst & Author By Tableau** | | **Apr 2020** |
| * Credential ID: b33c2571-fb44-4fe9-9cb7-9faa5ad76b25 ([**youracclaim.com/TableauAnalyst**](https://www.youracclaim.com/badges/b33c2571-fb44-4fe9-9cb7-9faa5ad76b25)) | | |
| **Tableau Desktop Certified Associate By Udemy** | | **Mar 2020** |
| * Credential ID: UC-50616b00-07ca-495e-9eb0-5cdb51afbd4d ([**www.udemy.com**](http://www.udemy.com)) | | |
|  | | |
| **Honors & Achievements** | | |
| **System Engineer** | | **Nov 2014 – Jul 2018** |
| Tata Consultancy Services Limited, Hyderabad, India | | |
| * Awarded “**Best Performer of the Month**” thrice and “**Best Performer of the Quarter**” once | | |
| **Student** | | **Jan 2019 – May 2019** |
| Illinois Institute of Technology, Chicago | | |
| * Runner up for presenting a paper on Modern Data Warehousing on Cloud Environments. Researched, explored and shared insights on the Cloud Data Warehouse processes like AWS Redshift, AWS Glue, Amazon Quicksight and their advantages over the traditional data warehouse. | | |