Samir Banerjee
Mobile: 09874666307

Professional Summary

- 11 years of total experience in the role of a developer, consultant and technical leader including three and a half-years-of experience as a data scientist and engineer using *Python*. I have worked in PSU, NBFC and Insurance verticals for both onshore and offshore clients. My core skillsets include SQL and Python and also I have working knowledge on Azure technologies like Data Factory, Databricks and PySpark.

- Good knowledge of SQL database complex Queries, Stored Procedures, Functions, Views and Triggers.

- Good knowledge of ETL / Data-warehousing architecture concepts including working knowledge on ADF.

- Good knowledge of OOPS concepts and *Data Structures* using *C++.*

- Good knowledge of Machine learning key concepts like Supervised, Unsupervised and Reinforcement Learning, algorithms like Linear Regression, Logistic Regression, KNN, Naive Bayes, Random Forest*,* etc. and creation, training and testing of data sets for descriptive and predictive analysis using Python libraries.

- Good knowledge of using Python on Dictionaries, Lists, Tuples, sets*,* for file handling and parsing of JSON objects.
- Strong technical skills and ability to lead and complete assignments within Scrum/Agile sprint cycles.

Summary of Projects

Independent Project
Name: Retail Analysis with Walmart

The aim of this project was to overcome the business challenge faced by Walmart due to unforeseen demands and out-of-stock issues. Walmart would like to predict the sales and demand accurately. Walmart runs several promotional markdown events throughout the year. These markdowns precede prominent holidays, the four largest of all, which are the Super Bowl, Labor Day, Thanksgiving, and Christmas.

The ML algorithm will predict demand accurately and ingest factors like economic conditions including CPI, Unemployment Index and the effects of markdowns on these holiday weeks.

Basic analytical tasks performed using Pandas and Matplotlib library included:

- To find out stores with maximum sales.
- To find out which store had maximum standard deviation i.e., the sales vary a lot. Also, find out the coefficient of mean to standard deviation.
- To find out which stores had good quarterly growth rate in Q3'2012.
- To find out holidays which have higher sales than the mean sales in non-holiday season for all stores together.
- To provide a monthly and semester view of sales in units.

For building, training and testing of Prediction Models *Scikit-Learn, Matplotlib and Statsmodel* libraries *were used.* Tasks included:

- Linear Regression – Analyze whether CPI, unemployment and fuel prices have any impact on sales.
- Change dates into days by creating new variable.

Project 1

Name of the Company: **<u>Capgemini Technology Services India Limited</u>**

Designation	: Senior consultant
Name of the project	: Chubb-2023-MDM-AIP SMS
Client	: Chubb Inc.
Domain	: Insurance
Platforms	: *MS-SQL Azure 12.0, Databricks, Cosmos DB, IBM DB2P,IICS, PyCharm , Github CI-CD.*
Project Duration	: January 2023- Present

- Worked in the *AIP Automation Insights* project where I did SQL enhancements based upon client requests logged in the Jira. Project delivery followed Scrum methodology where certain tickets were picked up by the team from the backlog list based upon client set criteria and resolved during the current Sprint cycle. Enhancements were implemented on the Views in the production database through *Git CI/CD* pipeline. Also worked on defects and implemented DML updates in the production database through Git CI/CD pipeline.
- Worked on Python part of the PRS Upgrade application and resolved P1 incident tickets related to Python scheduled script execution failure. *Python* scheduled scripts encrypt PII information sourced from multiple systems daily. Modified the script to change the encoding format that resolved the issue.
- Worked on *PRS C360 application* tickets logged in the Jira. Followed the best practices of ITIL and Scrum methodology to resolve the issues in the tickets created by the user.
- Worked on the automation of certain repetitive tasks such as table data purge activity for releasing disk space.

Project 2

Name of the Company: **Capgemini Technology Services India Limited**

Designation : Senior consultant
Name of the project : Chubb-MDM
Client : Chubb Inc.
Domain : Insurance
Platforms : *Microsoft SQL Azure 12.0,IBM DB2P,Databricks,CosmosDB,Postman.*
Project Duration : July 2022- December 2022

Project Summary

- ➢ The project related to ASP Customer 360 as the downstream application. Near real-time data was ingested from multiple upstream applications into SQL Datawarehouse through Kafka. Data was stored in the Stage tables and flowed to the Datamart tables after getting transformed as per business logic defined in the intermediary views. From Datamart tables data was uploaded to the driver table. The driver table data was picked up by the Databricks job and after further transformations got uploaded into the Azure Cosmos DB. From Cosmos DB data flowed through the API to the downstream ASP Customer 360 portal using Rest services.

Roles & Responsibilities

i) Technical leader of a team of four motivated professionals.

ii) Worked on the defects reported by the user and resolved them. Defects such as missing data, transformation and DDL logic change in SQL, debug Python job malfunction in Databricks to resolve malfunctions.

Defects were mainly related to customer payment updates, payment due date not visible in the portal, missing invoice, incorrect last payment amount & due date and customer policy related issues. Resolution was provided by checking the status through querying SQL table joins and querying the API JSON data through Postman. The status and the resolution were posted in the JIRA tickets created by user.

iii) Developed database level code fixes and tested in the UAT before deployment to the production database to correct defects.

Project 3

Name of the Company: **Capgemini Technology Services India Limited**

Designation : Senior consultant
Name of the project : Chubb Identity Management
Client : Chubb Inc.
Domain : Insurance
Platforms : Python 3.9,Visual Studio, Apache Kafka, Elasticsearch, Github.
Project Duration : April 2022- July 2022

Project Summary

➢ The project was in relation to creation and scheduling of a data ingestion pipeline from Kafka to Elasticsearch using Python. The data got written to the Elasticsearch staging Index and from there to the reporting Index which was then used for creating Kibana dashboards. Python was used for read-write operation and transformation of the data from Kafka to Elasticsearch. The project script was scheduled in the development server using Cron and is scheduled to run after every 10 minutes on a 24 * 7 basis.

Roles & Responsibilities

- Responsible for coding the data ingestion pipeline in Python from Apache Kafka to Elasticsearch as part of the Identity Management project.
- The pipeline extracted data from Apache Kafka to a staging index and then from the staging index to the reporting index in Elasticsearch. The data was then used in the Kibana dashboards for visualization.
- Wrote the pipeline code in Python and extracted raw data from Apache Kafka to the Elasticsearch staging index in the first stage. In the second stage the code applied transformations to the data as required before pushing into the reporting index.
- Uploaded the script, configuration and other parameter files in the Git repository and cloned the repository in Git Bash.
- Created a Cron job that is scheduled to execute the script after every 10 min and log the execution steps in a log file.
- Updated the Jira dashboard and provide updates to the scrum master on the day's work.

Project 4

Name of the Company:	Anunta Technology Management Services.
Designation	: Subject matter expert.
Name of the project	: Investment Management.
Client	: LTFS
Domain	: NBFC
Platforms	: Python 3.9, Anaconda 3, Jupyter, Oracle 11g, SQL Server 2012.
Project Duration	: 2020 – 2022.

Project Summary

➢ The project was related to Loan management and involved multiple ERP applications, EDW, cloud and on-premise databases.

Roles & Responsibilities

- Day-to-Day work involved data extraction and normalization, created of data frames and split into training and test datasets, built Regression models on the training dataset, validated the model using residual and outlier analysis, applied the model on the test

data to predict ROI, product wise sales, revenue growth, customer base, customer segment, market share and other KPI. Model used to create business rules and MIS dashboards.

Project 5

Name of the Company	:	Credentek Technologies.
Name of the project	:	Retail lending
Client	:	Family Credit
Domain	:	NBFC
Role	:	Database programmer.
Platforms	:	Oracle 11g, SQL Server 2012.
Project Duration	:	One year. (May 2016 – Mar 2017)

Project Summary

A lending project employed technologies like *Oracle*, *SQL,* and *Linux*. Project had several ERP applications implemented in a three-tier architecture.

Roles & Responsibilities

- Created query, package, *SQL* job and do debugging.
- Data load with *Oracle* loader.
- Tuned the queries and packages for optimal performance.
- Created scheduled jobs for generating daily reports.

Project 6

Name of the Company:		Protex Technologies.
Name of the project	:	Coalnet.
Client	:	Coal India Ltd.
Domain	:	ERP
Platforms	:	Oracle 10g, Sun Solaris, Java-script 6.0.
Project Duration	:	2012 - 2014

- ERP project provided data for MIS and day-to-day transactions to users across verticals in *Sales & Marketing, Finance, Personnel, Payroll and Stores Management*. Used *Oracle 10g* as the production database hosted in RAID configuration on a *Solaris* box *.JAVA/Java-script/JSP* used for implementing the UI and *Oracle Reports Builder* for hosting Reports in the application server.

Roles & Responsibilities

- Created queries on *Oracle* database for PIS (*Personnel Information System*) used for HR and payroll purposes. Data provided information to the client on employee salary, increment, leave, retirement, transfer, promotion, and exit.
- Took regular full-time backups of the MySQL production database.

Education

- Master of Computer Applications.

Personal Details

Name	: Samir Banerjee.
Date of Birth	: 08 July 1980.
Gender	: Male.
Marital Status	: Married.

Declaration

I hereby solemnly affirm that all the details furnished information is true to the best of my knowledge.

https://in.linkedin.com/in/samir-banerjee-08142425
https://r2d2.capgemini.com/IWRetainWeb/IWISAPIRedirect.dll/Files/static/html/webportal.html?page=talentprofile&resid=1631068

Place: Mumbai
Date: Samir Banerjee.