PARIN SHAH

San Francisco, CA| parin090199@gmail.com|(213) 285-8984 |linkedin.com/in/parin-s-shah|parin09.github.io/portfolio/|Tableau

EDUCATION

University of Southern California, Los Angeles   **Aug 2021-May 2023**
Master's in Computer Science (Data Science)   GPA: **4.0/4.0**

Mukesh Patel School of Technology Management and Engineering, NMIMS, Mumbai   **Aug 2016-May 2020**
Graduated as Bachelor of Computer Engineering with Distinction (Merit List)   **83.5%**

PROFESSIONAL EXPERIENCE

Unity Technologies, SF | Senior Data Engineer – Ads Team   **Jan 2023-Present**
- Leading product analytics, internal reporting and invoicing for the **Offerwall** product in the Ads org; partnering with PMs and business teams to deliver insights that drive product growth and support support strategic decision making.
- Building datasets by integrating third-party APIs (SensorTower, DeviceAtlas) to enrich feature platform inputs; improved data coverage by **30%** using fuzzy matching logic, and collaborated with DS teams to drive adoption in **ML models**.
- Designed and scaled a near real-time **OLAP** platform (Imply on Apache Druid), to enable **full-funnel analytics**—from ad requests to conversions to post-install events—supporting **400+** stakeholders with low-latency, high-concurrency querying.
- The OLAP platform enabled **70%** reduction in ad-hoc data pulls and cut insight turnaround time from **days to minutes** by democratizing access to granular funnel metrics for PMs and business teams to make faster data driven decisions.
- Migrated **50+** reports and dashboards from MicroStrategy to **Looker** by building aggregate tables and LookML models, ensuring data completeness and accuracy; reduced report load times by **40%** and seamlessly transitioned **100+** email subs.
- Optimized data pipelines and LookML codebase, reducing validation times by **4x** and saving **$50k** annually through code refactoring.
- Identified key areas for cost savings and optimizations by developing **GCP Cost Reporting** dashboard in Looker.

Unity Technologies, SF | Data Engineer Intern – Ads Team   **May 2022-Aug 2022**
- Discovered **1Mn+** new publishers using Unity to build games and advertising through Unity Ads by consolidating multiple datasets leading to an improved coverage over previous ETL from 2% to **65%.**
- Attributed a **63%** revenue increase to Unity-built publishers via in-depth analytics and enrichment.
- Technologies used: SQL, Python, GCP, BigQuery, Looker, LookML, Apache Airflow, Apache Druid, Github, Terraform, dbt.

Limechat, Mumbai | Data Scientist   **Jun 2020-Jun 2021**
- Reduced manual labeling time by over **60%** by devising a tool to label new intents using **clustering** and **active learning.**
- Increased product sales by **28%** on average across companies by incorporating results of **A/B** test to determine optimal number of products to be displayed.
- Designed an ETL pipeline in python to process and upload over **50,000+** rows of semi-structured data per day from PostgreSQL server to product analytics tool (**Mixpanel**).
- Developed a custom analytics dashboard on Flask and VueJS displaying critical **KPIs** for each client with ability to download standard excel and pdf reports. Integrated caching algorithm using Redis leading to **90%** faster loading times.
- Technologies used: Python, SQL, Mixpanel, Flask, VueJS, Azure, Postgres, Redis, Docker, Kubernetes, Grafana.

Oracle Financial Services Software, Mumbai | Research Intern   **May 2019-Jul 2019**
- Interlinked **LDAP** server in Kubernetes for authorization purposes. Automated manual formation of access roles in banking systems helping **save 10+** hours of work per client.
- Incorporated **OpenID** connect to fetch relevant roles from server. Authorized multiple users by assigning permission to roles with the help of **Role Based Access Control (RBAC)**.

TECHNICAL SKILLS

- Programming Languages and Databases: Python, SQL, SAS, Postgres, Redis, MySQL, Druid
- Python Libraries: Matplotlib, Plotly, Pandas, Numpy, Scipy, Tensorflow, Keras, Sci-Kit Learn
- Platforms and software: GCP, AWS, Azure, Looker, Tableau, Apache Airflow, Docker, Kubernetes, Imply, Git, Jira
- Big Data: Apache Spark, Hadoop, BigQuery, MapReduce, Kafka, dbt

ACADEMIC PROJECTS AND PAPER

Immersive Visualization in Medical Imaging: Reports 3D (Paper) | Python, Blender, Flask, VTK, Keras
- Constructed a web application to bridge the semantic gap between medical practitioners and laymen by leveraging the use of **Augmented Reality** (AR) as a graphically intensive solution.
- Devised novel algorithm for 3D **volumetric recreation** and **rendering** of organ. Generated 3D file to be viewed in AR.
- Achieved a **DICE** score of **87%** for segmenting HGGs (Brain Tumours) from MRI by applying U-Net CNN architecture.

Review of Credit Card Fraud Detection Techniques (Paper) | Python, Numpy, Sci-Kit Learn
- Implemented **Synthetic Minority Oversampling Technique** to generate synthetic dataset to improve size of minority class.
- Employed **logistic regression** to classify transactions and achieved a F1 score of **94%**, Recall of **99%** and Precision of **90%**.
- Analyzed currently existing credit card fraud detection algorithms based on various parameters such as pre-processing, complexity, computation time, accuracy and listed the advantages and disadvantages with a suitable use-case.

Recommendation System | Python, Spark, PySpark
- Designed a **hybrid recommender** system to predict the ratings by a user for a particular business on Yelp data.
- Achieved accuracy of **97.3%** by combining **item based collaborative filtering** and **model-based** recommendation systems.