#  Kanchan Choubey

 kr.choubey@gmail.com |408-475-1450 [|](https://www.linkedin.com/in/kanchan-choubey-52269a84/) SF Bay Area |[LinkedIn](https://www.linkedin.com/in/kanchanchoubey)|

**PROFESSIONAL SUMMARY**

* Around 6+ years of industry experience, worked with large datasets of structured, semi-structured and unstructured data for Analytics,

Business Analysis, Dashboards, ETL processes, Recommendation, Visualization, deploying data driven models and Reporting.

* Strong programming skills in Python, SQL, Tableau and Power BI and brief Experience in Project Management
* Master’s Degree in Business Analytics (Data Science)

**SKILLS**

 **Programming/Scripting/Cloud**: Python, R, SQL, Hive, Linux**/** IBM Cloud (Cloud Pak for Data), MS Azure, AWS(basic), Snowflake

**Database:** MySQL, MS SQL Server, MongoDB, SSMS, SSDT, MS Access, PostgreSQL, Microsoft Access

 **Big Data & Testing**  Hadoop MapReduce, Spark, Spark Streaming, A-B Testing, Multivariate Testing, ETL, Informatica, Ski-Kit Learn,

 & **Tools:** Anaconda, Jupyter

**Visualization/Reporting Tools**: Tableau, Power BI, Google Analytics, Kenshoo, Spotfire, DoubleClick, Social Media Analytics,MS Excel

**Management/Methodologies**: Jira, MS Project, AGILE, Waterfall, SDLC, MS Power Point, Azure DevOps, Zoho, Zendesk.

**Deployment Tools**: Amazon Sagemaker, Anaconda-Enterprise, Azure Machine Learning Studio,

#

**WORK EXPERIENCES**

#  Data Scientist-Analysis, Modelling & Reporting at Acutrack Inc.| Livermore, CA) July/20 – May/21

 ***Analytics:***

* Performed **data cleaning, data manipulation procedures** to optimize datasets for data analysis.
* Designed machine learning models for sales forecast and optimization of supply chain processes. These models ingested time series data based on various granularities to create algorithms using Pyspark modules.
* Designed, developed, and maintained Tableau functional reports based on Acutrack’s e-commerce and on-demand clients and also for
* internal team reports dashboard biweekly and monthly. Developed Tableau Dashboards for e-commerce data
* Reporting and analytical work with Tableau repository and performed regular ad-hoc reports in SQL Server.
* Designed and architected scalable data processing and analytics solutions and database Server replication, SSIS package and SSAS(Tabular Modeling).
* Build Optimization models to reduce the duration of section milling during the site closure. Minimized cost price by 40% providing the

optimized speed, torques and optimized other parameters with dimensionality reduction and feature engineering.

* Built ETL pipelines using Azure Data Factory (ADF) to create data flow from Azure Blob Storage to Azure Datalake Gen2 (ADLS)/Azure SQL Server and transformed data using Databricks.

 ***Project Management:***

* Work as a liaison between Production teams; ensuring timely delivery using **ManuOnline (ERP tool)** & **Zendesk**.
* Worked extensively on **Kanban Board** on Azure DevOps, and overseen project development and improvement
* Streamlined in liaising between business & technical areas with customer needs to achieve 98% customer satisfaction index.

#  Data Analyst- Pioneer Data Warehouse at Alumni Association| CSUEB, CA May/19 – May/20

* Maintained Alumni Database by **Tracking & Reporting** in **Salesforce CRM;** performedad-hoc tasksfor data of10k+ students using Python.
* Helped improve alumni engagement; created Alumni Dashboard depicting activity, statistics and demographics using MS Excel.
* Hands-on Experience in **MDX Expressions, DAX Expressions, Power BI Power Pivot**, Power integrated with Power BI. Data warehouse development and work with **SSIS and SSRS**.
* Created Alumni Dashboard on Tableau and designed a logistic regression model to predict engagement and gift generation and churn.
* Worked as **LinkedIn Account Manager** for CSUEB Alumni Page and generated monthly growth/engagement reports.
* Implemented Text Analytics and Sentiment analysis, creating word clouds, retrieving data from social networking platforms for Alumni
* Design, Develop and maintain Databases and improve extraction from the **Data Warehouse.**
* Improved the data gathering and extraction process with the **Pioneer Data Warehouse** with **optimized SQL queries** and by implementing mandatory data validation and implementing optimized methods to join tables**.**

**Data Analyst** at **Publicis Media| Mumbai**   **Sep/17 – Jan/19**

* Built and set-up paid ad campaigns across Search Engines like **Google, Bing, Facebook and Amazon**.
* **Optimized** campaigns as per set KPIs by Monitoring and Reporting, Impression Share, Quality Score, CTR, CPA, CPC, CPM, ROAS, ROI

conversion rate and changing keyword/placement bids and account daily/monthly budget caps.

* Improved conversion rates by over 40% and return on ad spends **(ROAS, ROI**) over 15% than the set benchmarks for some leading brands

and made weekly **strategic recommendations**.

* Performed exploratory data analysis, descriptive and inferential statistics on the data extracted from different sources and data types like structured (MySQL), Semi-structured (Json) and unstructured (Text files)
* Regularly work in system implementation projects (e.g. requirements documentation, systems configuration, test documentation/execution, issue identification and resolution).
* Conducted **Root Cause Analysis** for 1M rows of data collected from complex issues extracted using **MySQL** to recommend areas of improvement; resulted in a 35% boost in monthly performance and 30% depletion in resolution time
* Performed **A/B testing** to determine effectiveness of marketing campaigns to boost customer experience, to track user behavior for various clients’ mobile apps, compare it against desktop activities to make sure we deliver the right content to the right customer through right channel to reduce bounce rates and improve conversion by making sure that customers reach accurate landing pages.
* Design, develop, test, and deploy Power Query scripts in Power BI and perform detailed analytics for monthly reports.

**Data Analyst (Sales)** at **Circle Pro Audio| Mumbai**   **Dec/14 – Jul/1****7**

* Built SQL queries and reports to audit, maintain and improve weekly sales data of 8 categories of Audio Products.
* Generated automated reports in Power BI that reduced 50% of time execution to track sales insights.
* Designed plans with **Linear Programming Methods** and Decision Trees using **Solver Tables** and Precision Trees
* Generate sales trends, pricing, revenue performance metrics, product performance and other sales-related reports.
* Preprocess, Analyze, Sort, and process large amounts of revenue data using MS-Excel functions (**Power Pivots, V-lookups. H-Lookup**) and integrate to **Power BI**. Ad hoc report request reports and Create sub tables to create **sub reports**.

**Data Analyst** at **Carwale| Mumbai**   **May/13 – Oct/14**

* Defined the product roadmap & backlog to implement changes that drive user acquisition & user experience.
* Assisted in Lifecycle Management (PLM) through monthly market mapping and sizing of product capabilities.
* Generated **ad-hoc reports** using **SQL Server Reporting Services (SSRS)** to Improve the quality of the leads.
* Work closely with Carwale's stakeholders to design business models for product, sales operations, and sales teams
* Lead data projects and support systems projects creating transformational change through data migration, cleansing, and enrichment

**PROJECTS /DATA SCIENCE ENVIRONMENT/TECHNOLOGIES**

**Data Visualization Project-Power BI:** End-to-End BI Solution for main product- ‘bike’:

* Building a relational Data Model with (facts v/s lookup, star v/s snowflake) exploring best practices in Data Modeling and Power BI
* Add Calculated DAX Syntax and Operators to the Dashboard and Interactive Reports.

*Environment:*  Charts, Pivots, Formatting, Filtering, Connecting Tables, Analysis, Power BI

**Data Analytics Project-Python:** Analysis & Visualization of Job Market on portal “Simply Hired”**:**

* Used Beautiful Soup and Selenium to scrape online data.
* Selenium, NumPy, Pandas, Matplotlib, Geoplotlib, Seaborn libraries to scrape and visualize all available jobs and employment trends for

 Data Science in United States.

*Environment:* Beautiful Soup, Selenium, NumPy, Pandas, Matplotlib, Geoplotlib, Seaborn, Tableau, PowerPivot.

**Data Mining Project-R:** Predict telecom customers churn**:**

* Applied for Prediction and Prevention of Telecom Churn/Retention through Mining on Customer Data.
* Compared accuracy metrics, visualized and achieved 92% accuracy in classification tasks and studied the ROC curve

to study churn vary from 15% to 50 %.

 *Environment:* Data Cleaning, Data Preprocessing, Multicollinearity, Logistic Regression, Decision Trees, Random Forests,

Neural Network, Data sampling, Recommendation system, R, Dplyr, ggplot, gplot, heatmap

**Text Mining Project-Python:** Sentiment Analysis on Live Stream Tweetsusing NLTK **:**

* Created a program to live stream and analyze 200K+ Tweets using Tweeter API and Tweepy library.
* Explored correlation between Covid-19, economy and various other sectors using Sentiment Analysis.

*Environment:* Python, Tweeter API, TKinter, NLTK, Tweepy (StreamListener, OAuthHandler, Stream), Scikit-learn, Text-Blob, Text Mining,

Natural Language Processing,

**Machine Learning Project-Python :** Covid-19 Chat Bots:

* Scraped data from WHO FAQ’s and developed a retrieval based chatbot that answers Covid-19 related FAQ’s

 with 99% Accuracy.

 *Environment:* TensorFlow, pickle, nltk, Keras sequential API, Relu, Softmax, SGD, TKinter, WordNetLemmatizer, json, random, NumPy, Pandas.

**EDUCATION**

* **MS in Business Analytics (Data Science) |** California State University- East Bay (GPA: 3.31)
* **Bachelor of Engineering- Instrumentation|** Mumbai University