Mutkallam Warraich

wmutkallam@gmail.com | (814) 232-8252 | LinkedIn | GitHub

EDUCATION

The Pennsylvania State University **Expected Graduation: May 2027**
Bachelor of Science in Applied Data Sciences GPA: 3.47/4.00
Application Focus: Business Fundamentals

Relevant Coursework: Programming and Computation, Data Management, Data Privacy and Security, Data Integration and Fusion, Algorithmic Methods & Tools, Machine Learning for Data Analytics, Visual Analytics for Data Sciences, Programming Models for Big Data, Data Science Through Statistical Reasoning and Computation

TECHNICAL SKILLS

- **Programming Languages:** Python, R, SQL, Scala, Bash
- **Libraries & Data Technologies:** Pandas, NumPy, Scikit-learn, TensorFlow, Keras, Matplotlib, Seaborn, ggplot2, dplyr, PostgreSQL
- **Analytics & Development Tools:** Excel, Power BI, Tableau, Git, GitHub, Jupyter Notebook, pgAdmin
- **Data Science & Analytics:** Data Cleaning, Exploratory Data Analysis, Feature Engineering, Predictive Modeling, Statistical Modeling, Machine Learning, Deep Learning, Natural Language Processing (NLP), Hyperparameter Tuning, Model Evaluation, Data Visualization

EXPERIENCE

Strategy and Business Development Intern | H15DEN **Nov 2025 – Dec 2025**

- Collected, cleaned, and organized business and competitor data in Excel, developing structured datasets to support market segmentation, competitive analysis, and strategic decision-making.
- Analyzed 200+ businesses to identify high-value market opportunities and evaluate partnership potential, delivering data-driven recommendations to prioritize business development and targeted outreach.

PROJECTS

E-Commerce Customer Segmentation & Revenue Analysis **Aug 2026 – Sep 2026**
Tools: PostgreSQL, SQL, pgAdmin, Power BI, CTEs, Window Functions, Business Analytics

- Analyzed 5,000 e-commerce orders from 989 customers using PostgreSQL, applying CTEs, CASE statements, aggregations, and window functions to evaluate revenue trends, customer lifetime value, regional performance, and purchasing behavior.
- Segmented customers into High Value, Loyal, Regular, and Low Frequency groups and built a Power BI dashboard highlighting $5.11M in revenue, 950 repeat customers, and key customer, category, regional, and discount trends.

Product-Level Profitability & Revenue Optimization **Apr 2026 – May 2026**
Tools: R (tidyverse), ggplot2, dplyr, Business Analytics, Scenario Analysis

- Integrated and cleaned multiple e-commerce transactional datasets, engineering financial metrics including revenue, total cost, gross profit, and contribution margin to evaluate product and category-level profitability.
- Performed Pareto, margin, and scenario-based analyses to identify key profit drivers and high-cost transactions, finding that a 10% reduction in shipping costs could increase total profit by approximately 9.06% (~$217K).

Customer Churn Prediction and Risk-Based Retention Strategy **Feb 2026 – Mar 2026**
Tools: Python (Pandas, NumPy), Scikit-learn, Logistic Regression, Gradient Boosting, GridSearchCV, Matplotlib/Seaborn

- Built and tuned Logistic Regression and Gradient Boosting models to predict telecom customer churn, engineering tenure, pricing, and service engagement features, achieving a test ROC-AUC of 0.862 and capturing 66% of churners at a business-aligned threshold.
- Segmented customers into probability-based risk tiers, revealing a 73.7% observed churn rate in the high-risk group versus 12.5% in the low-risk group, informing targeted retention strategies and resource allocation.