PORTFOLIO DOCUMENT

Professor Ying Lin

Student Name: Aditya Pawar
SUID: 915517725
Email: aspawar@syr.edu
Portfolio GitHub link: github.com/adityapawarx/Ms-Applied-Data-Science-Portfolio
LinkedIn: https://www.linkedin.com/in/adityapawarx/

Project folders

Sr.no	Project Title	Link
1.	AI_travel_companion	GitHub - adityapawarx/AI_travel_companion
2.	A/B Testing for reward placement in a game using statistics	GitHub - adityapawarx/A-B_testing_statistics
3.	Marketing Mix Model	GitHub - adityapawarx/JulyPeakSavers-Forecasting-and-Mitigating-Energy-Surge-with-Data-Science
4.	Film Production Financial Analysis	GitHub - adityapawarx/Film-Production-Financial-Analysis

Introduction

Offered jointly by the School of Information Studies and the Martin J. Whitman School of Management, the Master of Applied Data Science degree program is designed to be a professional program of study, with a strong emphasis on the applications of data science to enterprise operations and processes, particularly in the areas of data capture, management, analysis and communication for decision making.

IST 688 Building Human Centered AI Apps

This hands-on course teaches techniques for developing and deploying intelligent applications powered by large language models. Students will gain technical skills in prompt engineering, knowledge enhancement, conversational interface design, and evaluating model outputs. Responsible AI development practices related to ethics, bias mitigation, and accessibility will be woven throughout. No prior AI experience required. Final projects provide an opportunity to apply concepts to build an AI system.

Project: AI_travel_companion

This project is a travel companion chatbot built using Streamlit and OpenAI's GPT-4o-mini, designed to provide an interactive and voice-driven experience for users. The application utilizes real-time geolocation data to detect the user's current location, retrieve weather conditions, and assist in travel planning, sightseeing recommendations, and language translation. It enables users to interact using audio input, and responses are provided in both text and speech formats using text-to-speech (TTS) technology. The chatbot supports multilingual conversations, ensuring that responses are given in the same language as the user's input unless they explicitly request translation. Additionally, the application allows users to upload images or take pictures via their device's camera, which are then analyzed to extract and translate text or generate relevant descriptions. The project also integrates vector-based search (ChromaDB) to retrieve relevant city attractions, shopping destinations, and upcoming events from a predefined dataset covering cities

like New York, Paris, Tokyo, and more. The chatbot follows a structured conversational flow, greeting users, confirming their travel plans, and providing personalized recommendations based on weather conditions and interests. Users can also scan and process PDF documents, making the application useful for referencing stored information dynamically. Overall, this AI-driven travel assistant offers a seamless, hands-free, and intelligent way to enhance travel experiences, making real-time recommendations and facilitating language translation while ensuring an engaging and user-friendly interface.

Through this project, I learned how to integrate AI-driven conversational agents with real-world functionalities like geolocation tracking, weather APIs, image processing, and multilingual support. Working with Streamlit and OpenAI's GPT-4o-mini, I gained hands-on experience in natural language processing (NLP), text-to-speech (TTS), and audio transcription, allowing me to build an interactive voice-based assistant. Implementing ChromaDB for vector-based search enhanced my understanding of embedding models and retrieval-augmented generation (RAG), enabling efficient retrieval of city-specific attractions and travel insights. Additionally, I deepened my skills in handling user inputs dynamically, including image uploads, PDF text extraction, and real-time chatbot interactions. This project reinforced my ability to design AI-powered applications that combine machine learning, API integrations, and user-friendly interfaces, while also highlighting the challenges of optimizing latency, model responses, and multilingual capabilities for a seamless user experience.

IST 652 Scripting for Data Analytics

The main content focus will be on information access and processing tasks on the types of structured, semi-structured and unstructured data in current use in information applications. For these three types of data, the course will include the use of structured numeric and text data such as that from a spread sheet or database, the use of data obtained through standard data exchange formats such as HTML or XML from web pages or JSON from web-based APIs, and the use of data obtained by pattern matching from text or log files. The

scripting language Python is chosen because of its ease of use and available packages to work with data in many information applications. The skills learned in this class are intended to complement the analytical and visualization skills learned in other data science courses. The scripting language Python will be taught, but it will be assumed that students already have a programming background, either through course work or through online study.

Project:
A/B_Testing_for_reward_placement_in_a_game_using_statistics

This project involves performing an A/B test for the mobile puzzle game Cookie Cats, developed by Tactile Entertainment, to determine the optimal placement of an in-game gate that affects player retention. The gate was initially positioned at level 30, requiring players to either wait or make an in-app purchase to continue playing. To test the impact of gate placement on user engagement, the game developers moved the gate to level 40 for a subset of players and compared 1-day and 7-day retention rates between the two groups. The dataset consists of 90,189 players, with each player randomly assigned to either the gate_30 (control group) or gate_40 (test group). The analysis focused on examining player behavior through game rounds played, calculating retention rates, and conducting bootstrapping simulations to measure statistical confidence in the observed differences. Results showed that retention rates were higher when the gate was placed at level 30, indicating that forcing a break earlier in the game helps sustain player engagement. The findings align with the concept of hedonic adaptation, which suggests that taking breaks from enjoyable activities prolongs engagement. The study concludes that keeping the gate at level 30 is the optimal strategy for maximizing player retention and sustaining long-term user engagement.

Through this project, I gained hands-on experience in A/B testing methodologies, statistical analysis, and player behavior analytics using Python and Pandas. I learned how to work with real-world experimental data, manage control and test groups, and apply bootstrapping techniques to measure the statistical significance of differences in retention rates. The project enhanced my ability to interpret user engagement metrics, recognize patterns in player

behavior, and understand how game design decisions impact retention and monetization. Additionally, I explored the concept of hedonic adaptation and its role in user engagement, reinforcing the importance of strategic friction in game mechanics to sustain long-term user interest. The experience helped me develop a structured approach to data-driven decision-making, emphasizing the critical role of experimentation and analysis in optimizing digital product experiences.

IST 707 Applied Machine Learning

This course will introduce the application of advanced machine learning methods for extracting patterns and knowledge from data to address stakeholder needs. The principles and theories underlying the methods will be discussed, and methods will be related to the issues in applying to real business and research problems. Students will acquire hands-on experience using state-of-the art tools to apply the methods covered. A central focus in this course is on communication, including translating stakeholder needs into technical methods, and communicating results in a transparent and compelling manner. The topics of the course will include the key tasks in applied machine learning, including data preparation, concept description, unsupervised and supervised techniques, deep learning, feature construction, optimization, evaluation, analysis, communication, and ethics. Through the exploration of the concepts and techniques of data analytics and practical exercises, students will develop skills that can be applied to business, science, or other organizational problems.

Project: Marketing_Mix_Model

This project focuses on Media Spend Cost Optimization using Marketing Mix Modeling (MMM) to analyze and improve the effectiveness of marketing channels such as TV, social media, digital ads, and traditional advertising. The goal is to measure the impact of media spending on product sales and optimize future budget allocations for maximum return on investment (ROI). By leveraging Bayesian statistical models and SHAP (Shapley Additive Explanations), we quantify the marginal contribution of each media channel, evaluate the ad-stock effect (lag in consumer response), and determine diminishing returns on media spending. The analysis helps companies allocate budgets more efficiently by identifying which channels drive sales and which

ones result in wasted spending. Key insights from this project include the realization that certain media channels have a delayed but lasting impact, some have immediate but short-lived effects, and beyond a certain budget threshold, additional spending does not necessarily increase sales. Using Bayesian MMM, feature importance analysis with SHAP, and optimization techniques, we create a data-driven budget allocation strategy that reduces costs while maintaining or increasing sales.

Through this project, I gained deep insights into Marketing Mix Modeling (MMM) and its real-world applications in media spend optimization. I learned how to apply Bayesian modeling techniques to analyze media impact, capturing factors such as ad-stock effects, diminishing returns, and seasonal trends. Implementing SHAP (Shapley Additive Explanations) improved my understanding of feature attribution in predictive models, allowing for a transparent breakdown of how each marketing channel contributes to sales. Additionally, working with real-world marketing data enhanced my skills in data preprocessing, feature engineering, and exploratory data analysis (EDA). I also developed expertise in budget optimization strategies using mathematical modeling, specifically applying SciPy optimization to maximize Return on Ad Spend (ROAS). Beyond technical skills, this project strengthened my ability to interpret business metrics, communicate data-driven insights to stakeholders, and align analytical models with strategic marketing decisions.

IST 659 Data Administration Concepts and Database Management

This is an introductory course in database management systems. It examines data structures, file organizations, concepts, and principles of database management systems(DBMS); as well as, data analysis, database design, data modeling, database management and database implementation. There is a specific emphasis on data analytics and learning to query data with Structured Query Language (SQL), query performance, data normalization; and database migration. This course provides hands-on experience in database design and implementation through assignments, lab exercises and course projects. This course also introduces advanced database concepts such as transaction management and concurrency control, distributed databases, multi-tier

client/server architectures, database applications, improving query performance through indexing, and advanced data query patterns for extract-transform-load.

Project: Film Production Financial Analysis

This project, Film Production Financial Analysis, focuses on applying data analysis techniques to solve key challenges in the movie pre-production phase. The primary goal is to help producers optimize budget allocation by analyzing expenses, actor and crew selection, and filming locations to ensure cost efficiency. Using SQL for data storage and querying, Power BI for data visualization, and reference datasets from IMDb and the American Film Institute (AFI), the project provides insights into cost trends across movie genres, the financial impact of casting decisions, and location-based cost variations. Interactive Power BI dashboards allow users to explore key metrics, such as budget distribution, revenue potential, and spending efficiency. By leveraging structured database analysis and visual storytelling, this project equips film industry stakeholders with data-driven insights to make more informed financial decisions during the early stages of movie production.

Through this project, I gained valuable experience in SQL-based data management, financial modeling for the film industry, and business intelligence visualization using Power BI. Working with structured film production data, I learned how to clean, store, and query datasets to derive actionable insights. Implementing interactive dashboards deepened my understanding of data visualization principles, allowing me to effectively communicate financial trends and patterns. Additionally, analyzing budget constraints, actor selection impact, and filming location expenses helped me develop a structured approach to cost optimization and scenario analysis. This project reinforced the importance of data-driven decision-making in entertainment finance, improving my ability to analyze real-world business problems, present insights visually, and assist in strategic budget planning for complex industries.
