National College of Ireland

Configuration Manual

MSc Research Project
MSc Data Analytics

Lakshmi Priya Kudupudi
Student ID: X23333367

School of Computing
National College of Ireland

Supervisor: Bharat Agarwal

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Lakshmi Priya Kudupudi
Student ID:	X23333367
Programme:	MSc Data Analytics
Year:	2025
Module:	MSc Research Project
Supervisor:	Bharat Agarwal
Submission Due Date:	11/08/2025
Project Title:	Configuration Manual
Word Count:	722
Page Count:	4

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Lakshmi Priya Kudupudi
Date:	15th September 2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Configuration Manual

Lakshmi Priya Kudupudi
X23333367

1 System Requirements

To run this project without any problem, I have taken the following setup:

1.1 Hardware Requirements:

Processor: Intel i5/i7/i11 or AMD processor.
RAM: 16 GB or 32 GB but I recommended 32 GB for large data
Storage: SSD with at least 256 GB space available
Internet: Stable broadband connection.

1.2 Software Requirements:

For operating system you must have Windows 11,Ubuntu 20.04 LTS or later.
Install Jupyter Notebook
Python Version: 3.12.0 or latest version
Must have Microsoft POWER BI desktop installed.
Create a Microsoft Azure account and enable blob storage containers.
Create a Salesforce developer org account.

1.3 Setting up Azure Account

First, I have created a Microsoft Azure account.
And then in Azure, I created a Storage Account and then a Blob Container to store my datasets and models.

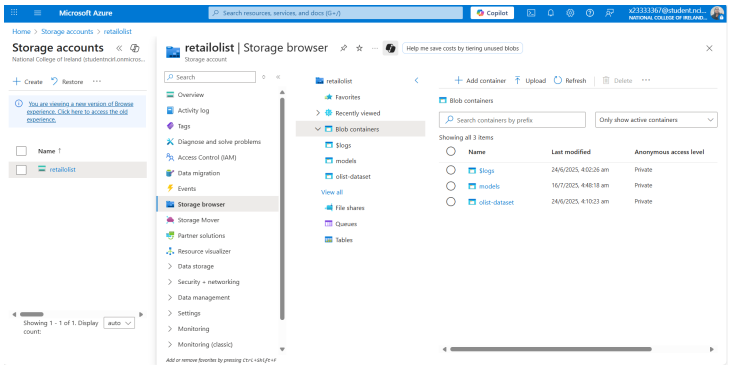

Figure 1: Azure storage and blob container

I downloaded the Olist public dataset from Kaggle and uploaded all CSV files to the Azure Blob Container. After uploading all the CSV files of datasets, I generated a SAS URL for secure access to these files and set the expiry date to September 15th, 2025. Attached the Azure Storage and blob container page for reference.

2 Import the Azure Data into Jupyter Notebook

I opened Jupyter Notebook on my local machine. Installed the Azure storage SDK and other required libraries.

Using the SAS URL, I loaded all datasets directly from Azure Blob Storage into Jupyter.I have checked all files to make sure data was complete and readable.

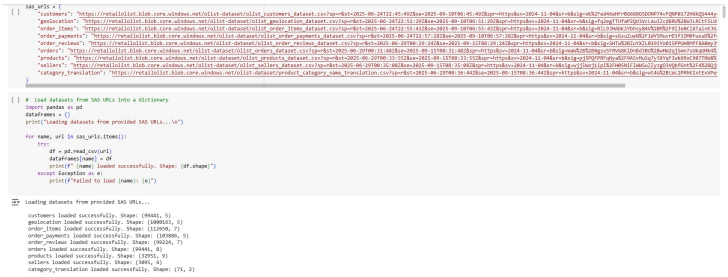

Figure 2: Imported Data from Azure to Jupyter

2.1 Data preprocessing

Removed missing values and duplicate records. Created new features like purchase frequency and average spend to improve model performance. Encoded categorical data and normalized numerical values. Split the data into training and testing sets.

2.2 Model Training

I trained four models for churn prediction:

XGBoost
Naive Bayes
KNN (K-Nearest Neighbors)
Bi-LSTM (Bidirectional LSTM)
All trained models were saved in Azure Blob Storage with name as models

2.3 Identified High-Risk Customers

From the predictions, I filtered only high churn risk customers, and selected only 2,000 records with the highest churn risk probability and uploaded it in salesforce. As we know that salesforce has storage limitations. So, that's why I have imported only 2k records into salesforce through code.

3 Salesforce Integration

Logged into Salesforce Developer Org. and Created a Custom Object named Retail_Churn_Pred__c. And then create required Custom Fields:

Customer_ID__c
Churn_Score__c
High_Risk__c (Boolean)

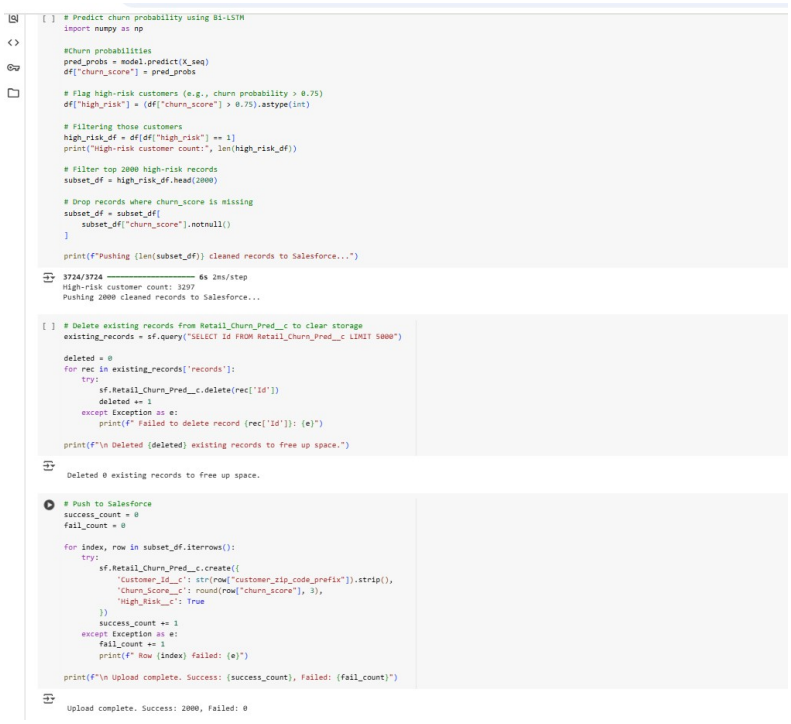

Figure 3: Salesforce Integration

Uploaded 2,000 high-risk churn customer records from Jupyter to Salesforce through the REST API. The Azure Blob connection was configured as per Salesforce guidelines Salesforce (2025). Deleted existing records and inserted new predictions for each execution.

4 Salesforce Automation

Updated Page Layout to show churn score and high-risk flag. As discussed in Dharmasena (2024), Salesforce workflows can be scaled using both declarative and programmatic approaches.

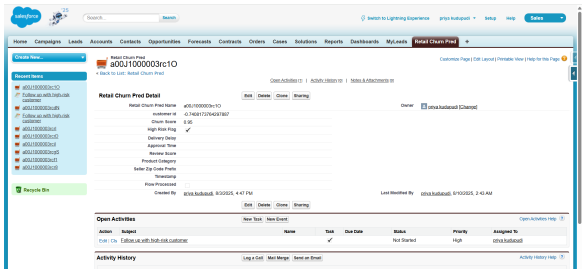

Figure 4: Salesforce Task Creation

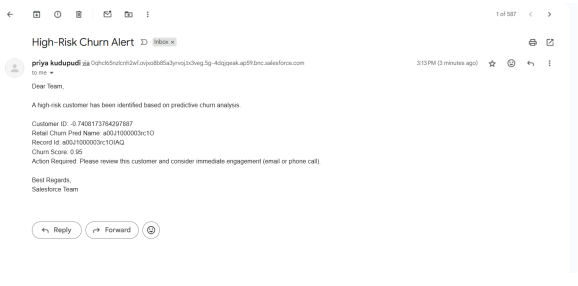

Figure 5: Salesforce Triggered Follow-up Email

Created a Salesforce Apex class and trigger to send a follow-up email and task to the CRM team when High_Risk__c becomes TRUE.

5 Power BI Dashboards

Open Power BI and refresh dashboards to see updated results. Connected Power BI to both Azure datasets and Salesforce churn prediction data. And then created dashboards for: Order Distribution by Status , Top 10 Seller Cities by Revenue , Monthly Cancelled Orders Trend , Top Product Categories by Number of Orders, Monthly Revenue Trend,Actual vs Estimated Delivery Time.The importance of churn prediction in retail has been well studied Patil et al. (2017), and our dashboard integrates these predictive insights into Salesforce for better decision-making.

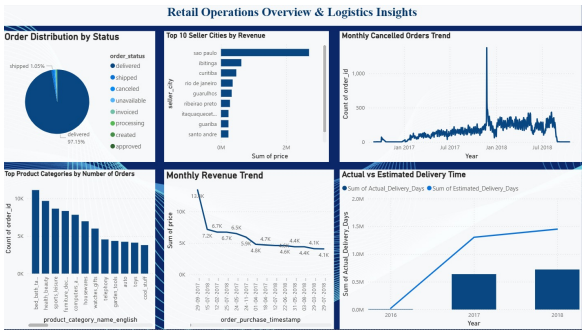

Figure 6: Retail Operations Overview & Logistics Insights

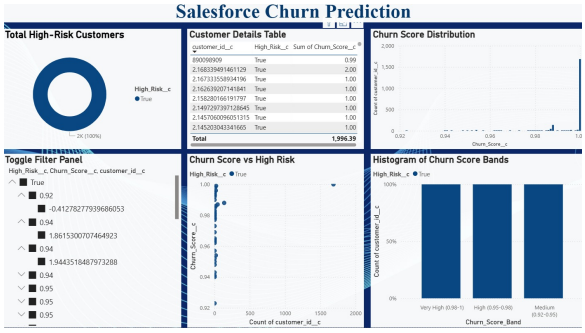

Figure 7: Salesforce Churn Prediction

Combined Azure retail data with Salesforce churn predictions for a full view.

Salesforce Churn Predictions:

Total High-Risk Customers,Customer Details Table ,Churn Score Distribution, Toggle Filter Panel ,Churn Score vs High Risk, Histogram of Churn Score Bands.

References

Dharmasena, S. D. (2024). Designing scalable salesforce workflows: Declarative and programmatic models in practice, Available Online.

Patil, A. P., Deepshika, M. P., Mittal, S., Shetty, S., Hiremath, S. S. and Patil, Y. E. (2017). Customer churn prediction for retail business, *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*, pp. 845–851.

Salesforce (2025). Setup an azure blob storage connection, `https://developer.salesforce.com/docs/data/data-cloud-int/guide/c360-a-set-up-azure-connection.html`.