TIKS ANALYTICS

EXTRACT |TRANSFORM |ANALYSE |SOLUTIONS

Major Research Project Proposal

FIRST DRAFT

CONTENTS

About Us

TIKS Analytics was formed in 2020 in response to the overwhelming demand for big data analysis. We are small but rapidly growing company that caters to the needs of various clients in North America and Europe. Our founders were team members at Georgian College who are today our partners within the industry. TIKS Analytics is a consulting firm specializing in the implementation and optimization of systems solutions. We focus on solving complex data analytics issues to enable our clients to make fact-based decision. We use sophisticated analytical tools and processes to uncover operational weaknesses, spot bottlenecks, and optimize performance in our client's business operations.

We extract and transform data using a combination of validated tools and techniques well known across the IT industry and within the field of big data. Our Business Intelligence (BI)suite includes several tools, methodologies and applications which help us collate large amounts of data in diverse formats from multiple sources. We create reports and dashboards to present results to our clients in a very easy to understand format. Business Intelligence when combined with analytics processes such as data mining, predictive analytics, and predictive modeling can provide greater insights, and identify new opportunities that can provide our clients with a greater competitive advantage.

OUR TEAM

Our team members are hardworking and dedicated professionals with their combined experiences spanning decades of skill and innovative thinking in manufacturing, consumer products, healthcare, financial services, and education bringing exceptional value to our clients.

Khadija Holder

BUSINESS ANALYST/TEAM LEAD
BSC. GEOGRAPHY; BSC MANAGEMENT STUDIES; PG DIP EDUCATION; MSC. GEOINFORMATICS
Pursuing a career as a Data Analyst to develop my problem-solving skills and to appease my constant desire to learn.

Tirthesh Jani

DEVELOPER
BSC. PHYSICS; PG DIP AIDI
Working towards a career as a research scientist in hopes of mastering and combining Artificial Intelligence and Physics.

Sukhwinder Singh

DEVELOPER
BSC. COMPUTER SCIENCE; MSC. COMPUTER SCIENCE
Pursuing a career in Big Data Analytics as an enthusiast of Data Science.

INTRODUCTION

CLIENT'S BACKGROUND

Falcon Logistics is a premier leader in logistics and supply chain management across multiple states including within the State of Iowa, USA. This company which was established in 1973 started out as a single outlet and gradually expanded to become a multi-state corporation. Our clients are contractors who deliver a wide-range of goods and services to their clients from small, specialized delivery packages to large batches of cargo.

CLIENT'S NEEDS

The logistics and supply chain management industry is a fiercely competitive one. Falcon Logistics is looking to be more competitive in incorporating alcoholic beverage vendors into their existing distribution market by gaining more contracts from alcohol vendors. Therefore, they would need to optimize their supply channels and delivery routes to incorporate new clients.

This industry is data-intense, and our client is looking for expertise to make decisions efficiently and effectively from large amounts of recent data. While they do some data processing in-house. Currently, their internal team is focused on the daily optimization tasks and needs. Therefore, they have decided to outsource this project.

CLIENT ACTIONS

By the end of the analysis, Falcon Logistics will have a deeper understanding of the alcohol market in the state of Iowa.

This will enable the client to make informed decisions and to take actions such as: -

- to control the flow of all the imported inventories
- to manage storage
- to conduct inventory management
- to devise marketing plans to target high demand clients

PROJECT DESCRIPTION

OBJECTIVE

The objective of this project is to investigate alcohol sales for logistics, supply chain and inventory management in Iowa, USA.

RATIONALE

TIKS Analytics role in this project is to extract and transform data for alcohol sales in Iowa, USA. The project will entail researching various data sources to gather reliable and valid datasets. We will assist our client by providing data insights towards attaining their goal of incorporating alcoholic beverage vendors in their existing market.

CLIENT'S CORE QUESTIONS

In a bid to increase their market share in Iowa, Falcon Logistics is looking to have the following questions answered.

- What are the recent trends in demand for alcohol sales in Iowa?
- Where are the retailers located?
- What is the annual demand for each retailer?
- How can we optimize our delivery routes to include potential high-demand retailers?
- How can we optimize our resources to be cost-effective?
- What are the future predictions for alcohol sales?
- How can we leverage our existing network to incorporate this opportunity?

RESEARCH QUESTIONS

The research questions are:

- To analyze the statistical significance of the sales in recent years in Iowa, USA
- To inquire into demand forecasting for alcohol sales based on seasonality for inventory management in Iowa, USA.
- To analyze the spatial distribution and its influence on alcohol sales in Iowa, USA.

- To identify high-demand alcohol products in various market segments in Iowa, USA (time of year, location, per capita)

SCOPE

This project is focused on: -

- Defining requirements for the project including key questions, SMART goals,
- Building a relational database
- Data cleaning to address data inconsistencies
- From the prospective of the products, the product analytics.
- From the prospective of the vendors, the inventory/demand analytics
- From the prospective of the liquor stores, the location analytics, and the demand analytics
- From the prospective of ML, Time series prediction technique, classification, clustering, and MBA.

OUT OF SCOPE

The following are out of the scope of this project: -

- Any recommendations made are not actionable for TIKS Analytics as they are out of the scope of this project.
- Changes will be managed within the scope of the project and any potential changes outside the scope of the project will be included in the recommendations for future projects to be pursued by the client.
- Maintenance of the IT infrastructure and data is out of the scope of this project.

OUTCOMES & BENEFITS

The client will be able to use the insights gained from this project to: -

- Strategically target potential clients in the alcoholic beverages industry.
- Better visualization of exiting routes and how there can be optimized.
- Visualization of potential routes for new high-demand clients.

DELIVERABLES

The project will be divided into three segments. Demand Forecasting, Spatial Analysis and Capital Margins.

The deliverables include: -

- Demand Forecasting: Historical Demand Trends, Demand Forecasts, Application Creation
- Spatial Analysis: Recent Demand Maps, Future Demand Maps, Comparative Maps
- Market Analysis: Sales Frequency, High Demand Products, Vendor Hierarchy
- A detailed written report
- Visualization of results
- Knowledge transfer of all research, workflows, analysis, visualizations, analysis models, testing, codes, formulas, algorithms, technology, documentation, and work product.

PROJECT EXECUTION

RESOURCES

The resources needed for this project include: -

- **Iowa Alcohol Demand Dataset**
  - The Iowa Department of Commerce requires that every store that sells alcohol in bottled form for off-the-premises consumption must hold a class "E" liquor license All alcoholic sales made by stores registered thusly with the Iowa Department of Commerce are logged in the Commerce department system, which is in turn published as open data by the State of Iowa.
  - This dataset contains 24 columns and 12 591 077 rows.

- **Iowa County Boundaries Feature Layer**
  - Iowa county boundaries service used for joining spreadsheet data to county polygons through ESRI Maps for Office. Contains two fields for joins (DOMCountyID and FIPS)
  - Contains Iowa DOM County Code (1-99) as a small integer, Census County FIPS as a both a string and integer.
  - This data was originally created by the Iowa DNR and digitized from USGS 7.5' topographic maps.

- **Access to Client's Database**
  - Falcon Logistics current logistics and supply chain database is essential in determining the up-to-date delivery routes, turnaround time, vehicle inventory, budget, sales, client list, volume of cargo, personnel and so on.

- **IT Infrastructure**
  - Computers
  - Internet Connectivity
  - Anaconda Package
  - ESRI ArcGIS Pro

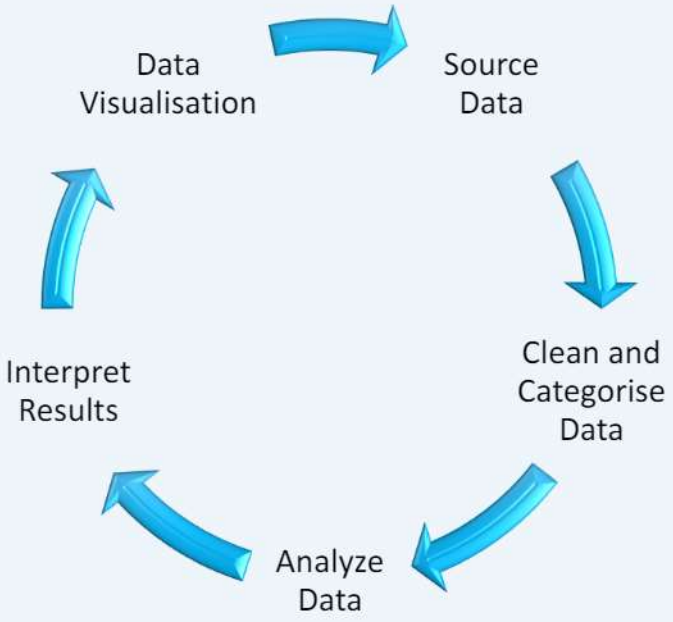
➲ **Data Dictionary**

- The data dictionary for the dataset outlines the data fields, data type, data length, description and constraints.

FIELD	DATA TYPE	DATA LENGTH	DESCRIPTION	CONSTRAINTS
IOWA ALCOHOL DEMAND DATASET				
Invoice/Item Number	object	20	Transaction invoice or item number	Not Null
Date	object	10	Date of transaction	Not Null
Store Number	int64	10	Number of stores that completed transaction	Not Null
Store Name	object	50	Name of store that completed transaction	Not Null
Address	object	100	Address of store that completed transaction	Not Null
City	object	50	City of store transaction	Not Null
Zip Code	object	5	Zip code of store transaction	Not Null
Store Location	object	100	Location of store that completed transaction	Not Null
County Number	float64	10	County number of store transaction	Not Null
County	object	50	County name of store transaction	Not Null
Category	object	10	Category designated by State	Not Null
Category Name	object	50	Category name designated by State	Not Null
Vendor Number	float64	10	Vendor number designated by State	Not Null
Vendor Name	object	50	Vendor name that completed transaction	Not Null
Item Number	int64	5	Item number of alcoholic product designated by State	Not Null

Item Description	object	50	Item description of alcoholic product designated by State	Not Null
Pack	int64	5	Number of packs	Not Null
Bottle Volume (ml)	int64	10	Bottle Volume of bottle in milliliters	Not Null
State Bottle Cost	object	10	Cost per bottle of alcoholic product designated by State	Not Null
State Bottle Retail	object	10	Retail Cost per bottle of alcoholic product designated by State	Not Null
Bottles Sold	int64	20	Number of bottles sold	Not Null
Sale (Dollars)	object	20	Sale of alcohol in dollars	Not Null
Volume Sold (Liters)	float64	20	Volume of alcohol sold in liters	Not Null
Volume Sold (Gallons)	float64	20	Volume of alcohol sold in gallons	Not Null
IOWA COUNTY BOUNDARIES FEATURE LAYER				
Perimeter	Number	20	Perimeter of county	Not Null
DOMCountyID	Number	5	County Identifier	Not Null
FIPS	Text	5	County FIPS	Not Null
Shape_Length	Number	10	Length of shapefile	Not Null
Shape_Area	Number	10	Area of shapefile	Not Null
FIPS (Integer)	Number	10	County FIPS as integer	Not Null
County Name	Text	50	Name of county	Not Null
State	Text	50	State that county is in	Not Null

DATA ANALYSIS PLAN



The following are the steps in the data analysis plan for this project:

🛠 Step 1- Source Data

The alcohol dataset was obtained from Iowa Department of Commerce and published by State of Iowa. It contains information on the name, kind, price, quantity, and location of sales, sales numbers of individual containers or packages of containers of alcoholic beverages. This dataset contains over 12 million rows and has a mix of data types all the way from strings to Boolean variables.

The county boundaries dataset was sourced from the State of Iowa through the Iowa county boundaries service used for joining spreadsheet data to county polygons through ESRI Maps for Office. It shows Iowa county boundaries comprising two fields for joins (DOMCountyID and FIPS), Iowa DOM County Code (1-99) as a small integer, and Census County FIPS as a both a string and an integer. This data was originally created by the Iowa DNR and digitized from USGS 7.5' topographic maps.

🛠 Step 2-Clean and Categorize Data

The alcohol dataset includes the store name, address, city, zip code, county, category, category name, vendor number, item number, item description, pack, size, state bottle cost, state bottle retail, bottles sold, sale dollars, volume sold (liters), and volume sold (gallons). The data would be cleaned and categorized by first removing any duplicates, then standardize the column headers, then create new columns for year, month, and day. Next the data would be categorized by store type (e.g. grocery store, liquor store, etc.), county, and item type (e.g. beer, wine, liquor, etc.).

The county boundaries dataset will be checked for null values and those will be treated appropriately. This dataset will be joined to a modified alcohol dataset.

🛠 Step 3- Analyze Data

The alcohol dataset can be used to train a machine learning model to predict the future sales of liquor in the state of Iowa. The machine learning model can be used to predict the sales of different types of liquor in different locations. The model can also be used to predict the sales of liquor in different price ranges.

The county boundaries dataset will be analyzed, using ArcGIS Pro software, to determine the spatial distribution of the alcohol sales market by county across the State of Iowa.

🛠 Step 4- Interpret Results

After analyzing the Iowa liquor sales dataset using machine learning, we can interpret the results in a few ways. People tend to buy more of one type of liquor than another, for example, people may buy more vodka than whiskey.

The results of the analysis of the county boundaries dataset will be interpreted, for example, there may be a correlation between the location of the liquor store and liquor sales, as people may repeatedly buy liquor from same store which can be the most convenient one.

🛠 Step 5- Data Visualization

After analyzing the Iowa liquor sales dataset using machine learning, these results can be visualized using various diagrams such as bar charts, which, for example, can depict the types of liquor on the x-axis and the total sales on the y-axis.

The results of the analysis of the county boundaries dataset will be visualized as tables, charts and maps using software such as ArcGIS Pro, MS Excel, and Power BI.

TOOLS & LIBRARIES

The tools and libraries to be used in the project include: -

- Anaconda
- ESRI ArcGIS Pro

- Jupyter Notebook
- Matplotlib

- Microsoft Excel
- Microsoft SQL Server Management Studio

- Notepad++
- Pandas

- Seaborn
- Sklearn

- Spyder

TECHNIQUES

The techniques to be used for each segment of the project are as follows: -

Demand Forecasting

Cleaning and Categorizing the Data

- The dataset that will be used for this application has 12 591 077 rows and 24 columns. It contains liquor sales in Iowa from December 2020 through to November 2021.
- The fields that will be used are city, invoice number, county name, store name, store number.
- A new column called 'liquor_type' will be added to the dataset to group the categories of liquor which is more user-friendly than the categories provided in the dataset.
- We will use machine learning to clean the data, remove outliers and sort the dataset.

Formulating Code

- Perform feature importance to find correct parameters.
- Code and train the models using the parameters.
- The resulting model should have some predictive data to be used for demand forecasting.

Creating Application

- The app consists of three sections - 'Insights', 'Exploratory Analysis' and 'Forecasting'. 'Insights' and 'Exploratory Analysis' contains data visualizations which provide powerful data-driven insights which can help the liquor companies with their manufacturing and distribution operations.
- The 'Forecasting' section provides a forecast of the future sales of liquor in the state of Iowa, the user is given the freedom to play around with certain parameters to tune the model and to generate a relevant forecasting model. The **Prophet library from Meta** will be used to generate the forecast and the trends.

Spatial Analysis

Steps in utilizing the Iowa dataset for spatial analysis of Iowa's alcohol sales demands.

1. **Clean the Dataset**- the dataset would be checked for null values, missing and/or incorrect data, misspelling and other errors.
2. **Categorize the Data**- the dataset would be categorized by vendor and the total demand for sales of each would be aggregated.
3. **Import the Data**- import the database into ArcPro software.
4. **Criteria for Sales Demand**- definition of criteria to evaluate demand, weighting of criteria, production of scenarios.
5. **Perform Spatial Analysis**- use spatial analysis tools to perform.
6. **Analysis of Results**- analyze results using an iterative method for optimal results.
7. **Generation of Recent Demand Maps**- generate demand maps showing recent trends of sales demand.
8. **Generation of Forecasted Demand Maps-** generate demand maps based on results of demand forecasting.
9. **Recommendation-** recommend suitable alcohol vendors to approach with service offers.

EXECUTION PROCESS

Software Development Life Cycle (SDLC)

Software Development Life Cycle (SDLC)

- Requirements
- Design
- Development
- Testing
- Release

TIME FRAME

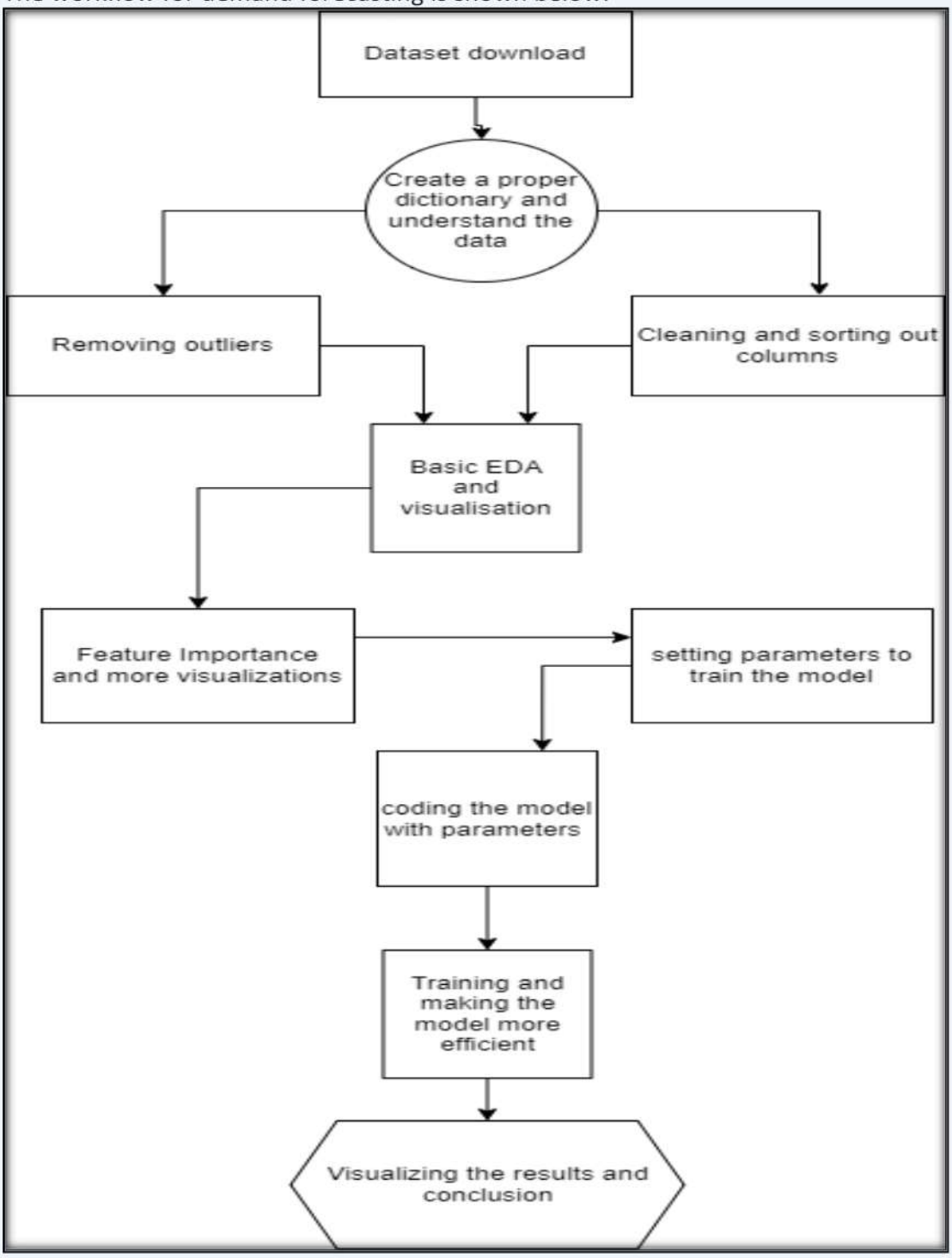
The project objectives and timeframe are shown below.

PROPOSED MAJOR RESEARCH PROJECT TIMEFRAME				
STEPS	**OBJECTIVES**	**MONTH**		
		1	2	3
Requirements	Define key problems			
	Refine research objectives			
	Seek client's approval			
Design	Acquire data			
	Outline techniques and tools			
Analysis	Aggregate data			
	Apply statistical instruments			
	Validate the results			
Testing	Analyse findings			
Release	Recommend the most feasible course of action based on the evaluation			

PROCESS VISUALIZATION

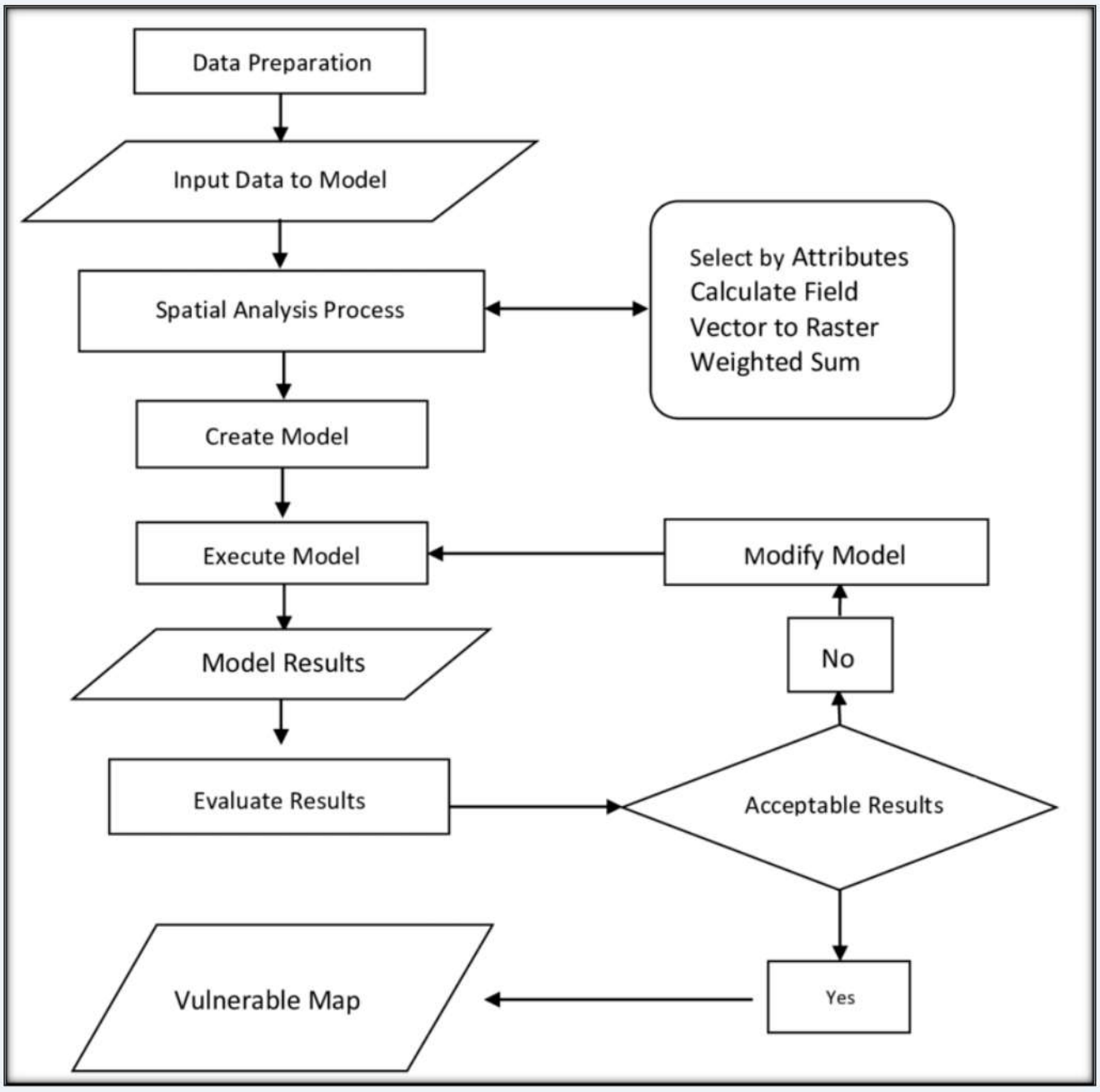
Demand Forecasting

The workflow for demand forecasting is shown below.



Spatial Analysis

The workflow for the spatial analysis is shown below.



POTENTIAL CHALLENGES

Whilst all future challenges cannot be predicted some of the more probable challenges include: -

→ **Project Delays**- can result from miscommunication and/or communication delays, illness, Acts of God, bureaucratic process (associated with government data).

→ **Scope Creep/Changes**- can result from the iterative processes involved in data analysis. These changes will be managed within the scope of the project and any potential changes outside the scope of the project will be included in the recommendations for future projects to be pursued by the client.

COMMUNICATIONS

The following is the communications plan for the project: -

NAME	ROLE	SEND/RECEIVE EMAIL	WEEKLY MEETINGS	DECISIONS	DOCUMENTATION
Falcon Logistics (Fake Client)	Stakeholder	Yes	Attend	Project Scope	Project documentation
Diego Carbajal	Instructor	Yes	Optional Attendee	Progression of Project	Grading
Brunilda Xhaferllari	Project Manager	Yes	Suggested Attendee	Project and Task Management related items	Project Management Presentation components
Khadija Holder	Team Lead	Yes	Attend- In-person & Online, Book, Minutes, Agenda	Team, Project, and Own Tasks	Communication- email, messaging, in-person
Tirthesh Jani	Team Member	Yes	Attend- In-person & Online	Project and Own Tasks	Communication- email, messaging, in-person
Sukhwinder Singh	Team Member	Yes	Attend- In-person & Online	Project and Own Tasks	Communication- email, messaging, in-person

CODE OF CONDUCT

ETHICAL CONSIDERATIONS

- We acknowledge our strict duty of confidentiality.
- We will ensure the observance of ethical practices demonstrating respect, honesty, and dignity.
- We will thoroughly follow the Client's directives without deviation from the objectives.
- Value is always present in our thinking.
- Validated data will solely form the basis of our observations, conclusions, recommendations, and decision-making.
- We undertake to use holistic analytics strategies and repeatable processes.
- We will thoroughly explain and document our analysis in detail for the Client. We will be transparent in all our activities.
- Moreover, we will check and recheck our research for quality, accuracy, completeness, and integrity before presentation to maintain the validity and credibility of the results.
- We will proactively circumvent unethical behaviour, such as exaggerating the results of our research.
- We will not falsely interpret, fabricate, embellish or otherwise misrepresent the data to validate our findings or change or omit details favouring making an analysis fit a hypothesis.
- We will be forthright and accountable if we make mistakes.
- We will tell the truth, even if it is bad news.
- Failure to observe deadlines is considered a breach of ethics.
- We have no conflict of interest and will not benefit from this Project.

INFORMATION MANAGEMENT

- We acknowledge that the data is an asset of any organization. The consequences and repercussions of unethical conduct when dealing with an organization's data can be significant and affect an organization's reputation, relationships and, ultimately, its revenues. Even the perception of unethical data handling has the power to undermine both internal and external trust.
- We give assurance that all the information provided is in our trust. It will remain private and protected from damage or alteration unless authorized.
- We will sign a Non-Disclosure Agreement that is acceptable to the Client and adhere to its conditions.
- We will notify the Client of the discovery of any sensitive information at any stage, namely personal identifying information or confidential information. We will make recommendations for addressing the issues and act according to their instructions.
- Original data and other information entrusted to us is stored in a secure location, such as SharePoint, and will remain unaltered.
- Copies of original data will be used for cleaning, discovery, manipulation, and analysis.
- Naming conventions and version controls will segregate documentation, datasets and work product. Each will align with the applicable phase of the Project.
- All work will only be conducted through secure applications.

- All our workstations are protected by a password that is unknown to anyone other than the assigned user. No one has access to our workstations, including colleagues, family members and friends.
- All information sharing is secured from potential risks using encrypted channels, such as SharePoint.
- Discussions about the data, analysis and observations will only be amongst client-authorized collaborators and our course instructors.
- All information entrusted to us will only be used for its intended purposes unless specifically approved by the Client.

PROJECT MILESTONES

The following are the project milestones.

MILESTONE	COMPLETION DATE
Acceptance of Project Charter	Day 2
Requirements	Week 1
Design	Week 2
Development	Week 5
Testing	Week 8
Release	Week 9
Presentation	Week 10
Handover of the Deliverables	Week 12
Project Close-Out	Week 12