Smart Cities: *Infrastructure and Transport Sector Competitions*

1st Aseel Alsaeed
Management Information System
King Saud University
Riyadh, Saudi Arabia
443202167@student.ksu.edu.sa

2nd Sarah Alotaibi – Group Leader
Management Information System
King Saud University
Riyadh, Saudi Arabia
443201865@student.ksu.edu.sa

3rd Sara Almawash
Management Information System
King Saud University
Riyadh, Saudi Arabia
443201064@student.ksu.edu.sa

4thMeznah Alsubaie
Management Information system
King Saud University
Riyadh, Saudi Arabia
443200846@student.ksu.edu.sa

5th Aljoharah Alsulami
Management Information System
King Saud University
Riyadh, Saudi Arabia
443202090@student.ksu.edu.sa

6th Sadeem Alzahrani
Management Information System
King Saud University
Riyadh, Saudi Arabia
443200567@student.ksu.edu.sa

I. INTRODUCTION

As part of Saudi Arabia's Vision 2030, which, as highlighted on its official platform, identifies smart cities as a cornerstone for achieving sustainable development and improving quality of life [1], this project seeks to investigate the scope and nature of investments made by various stakeholders in this domain, while identifying existing shortcomings. Furthermore, evidence provided by the Saudi government's Open Data Platform underscores the growing attention directed toward the development of smart cities [2], thereby necessitating greater efficiency in investment practices and a more explicit alignment with the overarching objectives of the Vision.

II. BUSNIESS PROBLEM

Saudi Arabia is investing heavily in infrastructure projects to support its goal of transforming all major cities into smart cities. Hundreds of infrastructure tenders are issued each year by agencies with a vested interest in infrastructure, and government authorities select the projects they choose to fund. Despite the scale of this activity, there is limited transparency and analysis of how budgets and timelines are allocated across authorities and projects. Some projects require large, long-term investments, while others can be delivered quickly and at lower cost, yet these differences are not evaluated systematically.

Therefore, there is a need to analyze tender patterns to avoid inefficient budgeting and unrealistic timelines. Our project will examine 2024 tender data to uncover trends, patterns, and outliers in cost and duration, providing actionable insights to help allocate budgets more effectively and manage project schedules more accurately.

III. DESCRIPTION OF THE DATASET AND THE ATTRIBUTES

Dataset Description: The dataset titled "Infrastructure and Transport Sector Competitions" is stored in Microsoft Excel format (.xlsx). It consists of records detailing government competitions (tenders/projects) within the infrastructure and transport sector. The dataset provides valuable insights into the nature of these competitions, covering their scope, responsible entities, financial values, and timeline information.

The dataset contains 1312 entries(rows) and 10 columns, each representing a unique competition in the infrastructure and transport domain. This allows for analysis of the sector activities, funding allocation, execution timelines, and decision-making processes.

TABLE I. DATASET ATTRIBUTE

COLUMN NAME	DATA TYPE	DESCRIPTION
YEAR	INTEGER	THE YEAR IN WHICH THE COMPETITION WAS STARTED
COMPETITION NAME	CHARECTER	THE NAME OF COMPETITION.
COMPETITION NUMBER	INTEGER	ID NUMBER FOR COMPETITION
DESCRIPTION	CHARECTER	ADDITIONAL DETALS ABOUT THE COMPETITION.
SECTOR	CHARECTER	THE SECTOR UNDER WHICH COMPETITION FALLS.
APPROVING AUTHORITY	CHARECTER	THE AUTHORITY RESPONSIBALE FOR APPROVING THE COMPETITION
START DATE	INTEGER	THE DATE WHEN THE COMPETITION STARTED.
END DATE	INTEGER	THE DATE WHEN THE COMPETITION ENDED.

COMPETITION STATUS	CHARECTER	THE STATE OF THE COMPETITION.
VALUE (IN RIYALS)	FLOAT	THE COST OF THE COMPETITION IN RIYALS.

IV. SELECTED DATA MINING TECHNIQUES

To address the business problem, we will apply both clustering and regression techniques in order to uncover patterns in tender data and to predict project budgets and timelines.

Regression can be used to estimate project budgets and durations by analyzing relationships between project attributes such as competition name, approving authority, start date, and end date. For example, if past data shows that tenders from a specific authority with longer durations tend to require higher budgets, regression can provide a numerical estimate for similar new tenders. This prediction supports better financial planning and more accurate scheduling by forecasting the likely cost and duration of upcoming projects.

Clustering, unlike regression, does not predict a specific value but instead groups tenders into categories based on shared characteristics. In our dataset, clustering can be applied using attributes such as project budget and project duration. This allows us to identify natural groupings such as low-cost short-term projects, and high-cost long-term projects.

By integrating regression and clustering, actionable insights can be derived from the tender dataset. Together, these techniques enhance the ability to analyze tender data, highlight inefficiencies, and support strategic decisions that align with Saudi Arabia's smart city initiatives.

[1] Kingdom of Saudi Arabia, Vision 2030, [Online]. Available: https://www.vision2030.gov.sa

[2] Saudi Open Data Platform, Infrastructure and transport sector competitions, [Online]. Available: https://open.data.gov.sa/en/datasets/view/50e49731-2d9f-453f-b20d-c43c57117673