Database Design Project

John J. Bolling
Business School, University of Exeter

BEM2040: Database Technologies for Business Analytics

28.03.2024

Word Count:

This project will include the following.

• Use case scenario which has a business problem that this database will look to solve
• Entity-Relation Diagram
• Normalized form in 3NF
• Implementation into SQL
• Python to extract

Use Case Scenario

Business Problem

Southwest Water, a renewable energy management company, owns a series of wind turbines in Devon. Due to recent uncertain weather conditions and the reactive nature of the company's managers and plant operators, the Den Brook Wind Farm in Devon has been operating at a loss. Running a wind turbine plant is costly, so the company needs a profit-making plan to continue its operations.

Jordan is a data analyst at White Space Strategy who has been tasked with tackling this issue. He has planned to construct an application to minimize operational costs by achieving the following.

• Minimize operational costs: Create a short-term plan that optimizes wind energy production and reduces inefficiencies in operations.

• Suggest long-term investment options to improve productivity.

Application Overview:

• User logs into the application using credentials (Username and password).

• Initially views ‘Wind Data’ and then ‘Wind Power Data’ to identify when the winds are highest and lowest to see if there is a correlation between energy produced and independent variables such as wind speed, turbine direction and turbine condition, which days had volatile weather conditions and if absences have led to operational inefficiency causing a difference between expected and actual energy produced.

• He then uses the ‘Predictor Model’ and ‘Prediction Result’ to analyze forecasts for wind speed, direction and other relevant weather conditions to generate to construct his business plan which will be evaluated later.

Further Details of Metrics Used in the Application:

• Absences: 2 Absences tables for week 11 and 12. ‘W11Absences’ are the number of absences during the week 2024/03/11 to 2024/03/17. ‘W12Absences’ refers to week 2024/03/18 to 2024/03/24. ‘DaysAbsent’ is a record of which days the user was absent and ‘AbsencesInfomation’ is the reason of absence.

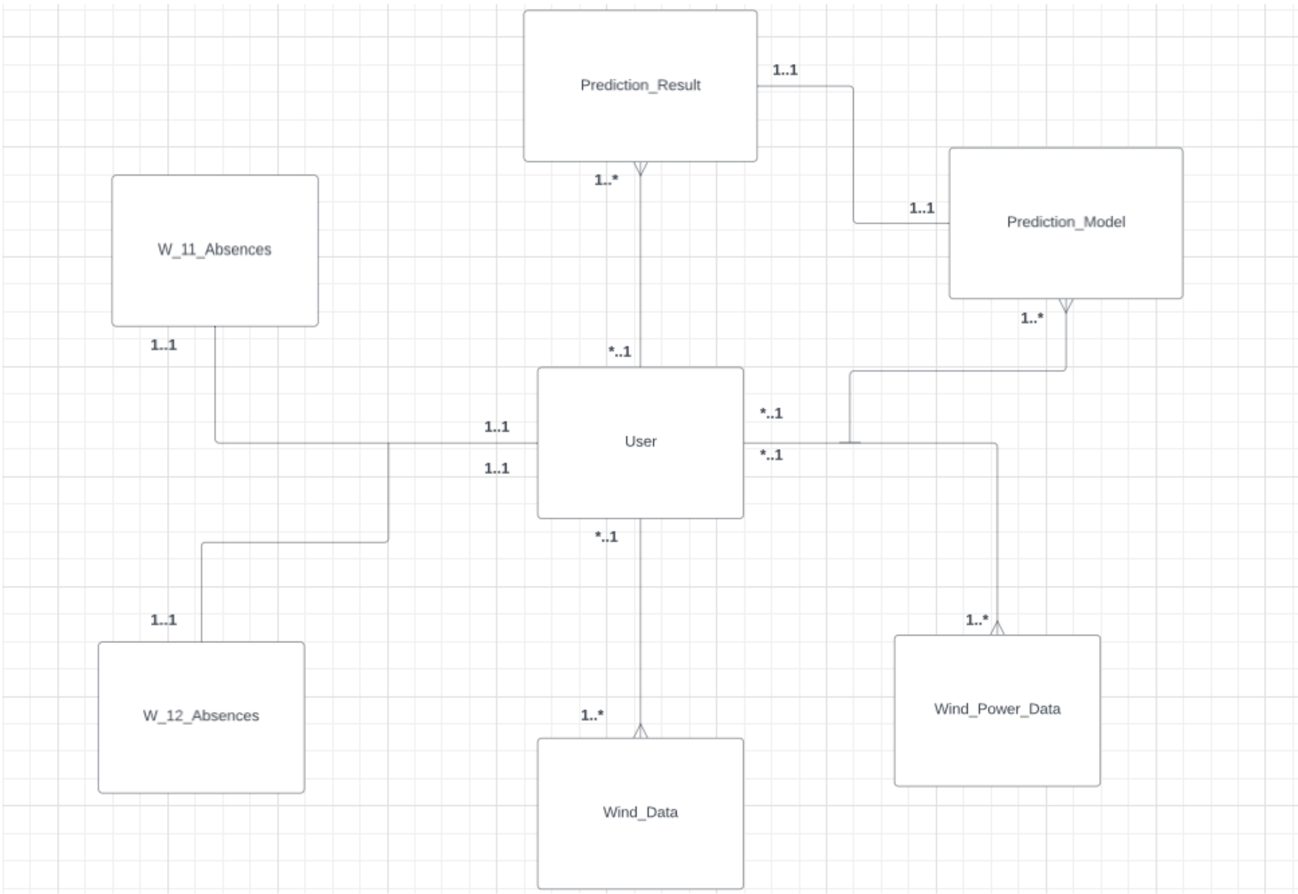
• Wind Data: ‘Timemarker’ is a time stamp, and the wind data is recorded every 8 hours. Wind speed (measured by an anemometer, in ‘mph’) and wind direction (measured by a wind vane in degrees, where 0 indicates north and 90 indicates east), which is recorded at a particular time within a day, are recorded at the time stamps. Additional weather information is also noted. All the above data is sent by the nearest weather station and is then entered manually onto the database.

• Wind Power Data: Expected and Actual energy produced, which considers wind power generated from the turbines, is measured in Watts(W) by a generator that converts the rotational energy of the turbine into electricity. 'TurbineInfomation' records the condition of the turbine to see if it is running optimally. Expected and Actual turbine direction is also recorded in degrees (same metric as 'WindDirection'). Here, the blades are perpendicular to the wind direction. The same time stamp in the wind data is used here.

• Prediction Model: Has the timestamps, 'TimeOfPrediction', that is recorded every 12 hours. The model utilizes predictive analytics tools to forecast the metrics used in 'PredictionResult'.

• Prediction Result: The same time stamps as 'PredictionModel' are used and the forecasted wind speed and direction, using the same metrics as above are recorded. Additional information, similar to 'WindData' is included as well.

Entity-Relationship Diagram



Relationships

• User-Absences: One-to-one, each user has their own absence record.

• User-Wind and wind power data: One-to-many. One user can create and manipulate any data.

• User-Prediction model and result: One-to-many. One user can create and manipulate any data.

Data Dictionary in 3NF

The following tables contain an example of an entry and the number of entries.

Users table:

UserID (PK)	Username	LoginPassword
1	john_doe	P@ssw0rd1

10 rows (Jordan, 1 manager and 8 plant operators)

W11Absences table:

AbsencesID	Username (FK)	W11Absences	DaysAbsent	AbsencesInfomation
90	john_doe	0	N/A	No Absences

10 rows for each user

W12Absences table:

AbsencesID	Username (FK)	W12Absences	DaysAbsent	AbsencesInfomation
90	john_doe	0	N/A	No Absences

10 rows for each user

Wind Data table:

WindDataID (PK)	Timemarker	WindSpeed	WeatherInfomation	WindDirection
1	2024-03-13	14 mph	Sunny	0

30 rows

Wind Power Data table:

WindPowerDataID	**Timemarker**	**ExpectedEnergyProduction**	**ActualEnergy Production**	**TurbineInfomation**
1	2024-03-13	280 W	280 W	Optimal condition
ExpectedTurbineDirection	**ActualTurbine Direction**			
0	0			

30 rows

Prediction Model table:

PredictionModelID	**PredictionReInfomation**	**TimeOfPrediction**
1	Future Analysis	2024-03-23 00:00:00

20 rows

Prediction Results table:

PredictionResultID	**TimeOfPrediction**	**WindSpeedPrediction**	**WindDirection Prediction**	**ResultNotes**
1	2024-03-23 00:00:00	10 mph	0	Clear

20 rows

These tables are in,

1NF: No multi-valued attributes in any column, each column name is unique and there are no repeating groups.

2NF: No composite primary keys

3NF: Values in the non-primary key column can only be determined from only the primary key.

Report

This report will include an evaluation and reflection.

Evaluation

The evaluation will consider Jordan's initial analysis, how he will overcome challenges faced in his analysis and the business outcome.

Analysis

Jordan looks at the wind and wind power data and establishes the analysis below.

When turbines were in the optimal direction and condition respectively	**Energy produced (W)**
Highest wind speed: 30 mph Time stamp: 2024-03-15 16:00:00	600
Lowest wind speed: 5mph Time stamp: 2024-03-18 08:00:00	100

When turbines were in optimal condition but in suboptimal direction	**Energy produced (W)**

Wind speed: 25 mph Wind direction: 55 Degrees Turbine direction: 80 Degrees Time stamp: 2024-03-20 16:00:00	Expected: 500 Actual: 425
When turbines were in optimal direction but in suboptimal condition Wind speed: 16 mph Turbine condition: Suboptimal Time stamp: 2024-03-21 08:00:00	Expected: 320 Actual: 160
Wind speed: 18 mph Turbine condition: Suboptimal Time stamp: 2024-03-21 08:00:00	Expected: 360 Actual: 180

The correlation between Wind Speed and Energy appears to be a strong positive.

The mean wind speed from the wind data was calculated using the python code provided using numpy and from scratch, which is 15.13 mph (2 d.p). As this correlates strongly with energy produced (1 mph approximately equals 20 W generated), Jordan can use this to predict the energy produced as 302.6 W. This is useful in Southwest winds budgeting processes and comparing expected revenue with budgeted total cost.

If a change of 10mph or 60 degrees occurs between 2 consecutive time stamps, then that day is considered a volatile one. Considering this, 20th and 17th of March 2024 were volatile, which was 2 out of the 10 days considered.

Considering the data from the prediction results, Jordan presented the following issues generated from past data and long-term investment options to solve these problems and improve productivity.

• Hire more workers: The increase in absences in W12 led to the misplacement of turbine direction in the time stamp, 2024-03-20 16:00:00. This day was volatile and the workers on the day couldn't handle the fluctuations. There were 3 absences on the 21st of March, which was a day when the turbines required maintenance. This led to a massive drop in energy produced. Operator Laura Miller has taken maternity leave and will need replacing as well, which increases the need for more workers.

• Invest in innovation: Through this, Southwest Winds can research and develop trackers to monitor turbine health as well as automatic turbine direction changers that respond to wind direction changes. If successful innovation is achieved, the above-mentioned inefficiencies wouldn't occur, and the plant wouldn't need as many operators, which could reduce labor costs.

The following recommendations were made to minimize operational costs in the short term.

Day	**Recommendation**
2024-03-23	Adjust blades to 0 and 10 degrees at 12 am and 12pm respectively. Consistent day so additional workers aren't needed.
2024-03-24	Adjust turbine to 10 and 20 degrees at 12 am and 12pm respectively. Consistent day so additional workers aren't needed.
2024-03-25	Adjust turbine to 45 and 60 degrees at 12 am and 12pm respectively. Have more workers on this day as a thunderstorm causes a big fluctuation in wind speed. Keep more contact with weather stations.
2024-03-26	Adjust turbine to 70 and 5 degrees at 12 am and 12pm respectively. Have more workers on this day as the thunderstorm

	causes another big fluctuation in wind speed. Keep more contact with weather stations
2024-03-27	Adjust turbine to 10 and 15 degrees at 12 am and 12pm respectively. Consistent day so additional workers aren’t needed.
2024-03-28	Adjust turbine to 15 and 20 degrees at 12 am and 12pm respectively. Consistent day so additional workers aren’t needed.
2024-03-29	Adjust turbine to 40 and 35 degrees at 12 am and 12pm respectively. Consistent day so additional workers aren’t needed.
2024-03-30	Adjust turbine to 20 and 200 degrees at 12 am and 12pm respectively. Have more workers on this day as heavy rain causes a big fluctuation in wind direction. Keep more contact with weather stations.
2024-03-31	Adjust turbine to 100 and 0 degrees at 12 am and 12pm respectively. Have more workers on this day the heavy rain causes another big fluctuation in wind direction. Keep more contact with weather stations.
2024-04-01	Adjust turbine to 0 and 10 degrees at 12 am and 12pm respectively.

This plan will reduce the plant’s cost per unit, consisting of maintenance cost and wages, of electricity produced. The minimization of these costs would improve the branch’s profit and ROI. The plant and company’s competitive advantage in the renewable energy industry can be strengthened while achieving sustainability goals.

Reflection

Workshop 1

In this workshop we had to develop an idea for a use case scenario and a database that can be created to achieve the business problem identified in the use case scenario. The ERD design and normalization process, to be implemented once the use case scenario is refined, was also explored. We were able to select from many examples of coursework options provided in the module. We could take a management (ex: inventory management system), an economics (ex: financial portfolio management system) or a sustainability (ex: renewable energy management) focus. My use case scenario is a renewable energy management approach that includes management-style cost cutting and employee focus as well.

Workshop 2

In this workshop, we presented our ERDs for the project, which was critiqued. We also grasped the concept of entities, attributes and keys in databases, and the normalization process. It was only after this workshop that I understood how to convert my unnormalized form into a 1^{st} ,2^{nd} and 3^{rd} normalized form, which I have presented in this report.

Workshop 3

In this workshop, normalized forms in 3NF were presented to receive feedback on and the implementation into SQL and, the extraction and analysis on Python was briefly discussed. I was able to fully understand and refine the business logic of my project and adjust my ERD, normalized form and subsequent SQL database creation.

After further adjusting my database, making additions and omissions, to formulate a more well-rounded business problem and solution, I was left with only python extraction and analysis to tackle.

Implementation Into SQL

I varied the amount of dummy data for each table appropriately as discussed in the normalized table section to maintain the realism of the model and successfully implemented ran the code.

Extraction and Mean calculation in Python

Utilizing the IT services of the university, I installed miniconda (W1 setup instructions document link provided) and checked the integrated python version (Command prompt: python –version). I then installed and successfully ran the pyodbc (Command prompt: conda install pyodbc) and numpy files (Command prompt: conda install numpy). This enabled me to successfully achieve the mean value of wind speeds from wind data to use for analysis.