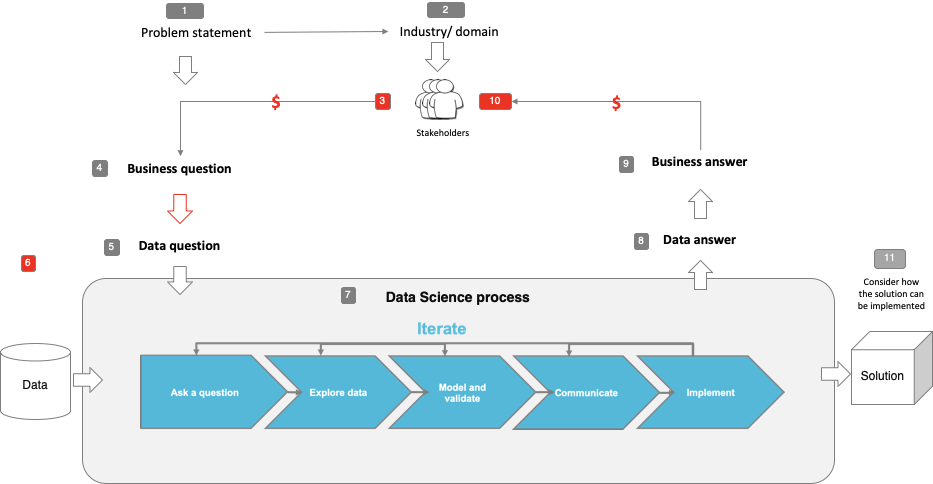
**Diego Molina**

**Capstone Project Document**

# Process overview

The following diagram shows the overall end-to-end process for defining, designing and delivering the Capstone project.



# Problem statement

Ozwide tools are facing a quick growth in sales per year, around 1.5 million was the total turnover last financial year which is the highest turnover amount the company has ever recorded in its history.

**Why is it important to address this problem?**

During this year Ozwide Tools had to order products using air freights which are around 30% more expensive than sea freights, increasing the purchase costs and decreasing gross profits.

The company is looking at substantially reducing these costs by having a reliable forecasting system that fills the market needs and saves money on freight costs.

**What is the current state?**

Ozwide Tools is ordering at least 20% of the orders using air freight rates losing a 5% to 15% profit due to the extra expenses.

**What is the desired state?**

Ozwide Tools is looking to make a scheduled ordering program based on demand forecast to order on a 2-month distribution using just sea freights and reducing the need of air freights by 10%.

**Has this problem been addressed by other research projects? What were the outcomes?**

Ozwide Tools is using a naive model based on average sales per day to forecast the demand. We are looking in improving this system to get better results

# Industry/ domain

Ozwide Tools is an automotive speciality tools provider which is a niche and very specific field in the Automotive Industry. On the sales side, big companies such as Burson, Repco, Hsy and Cool Drive are in play however, on the other side which is the rentals Ozwide Tools is a sole player which gives the company a strong position in the market.

Our company is highly related with the spare parts business which is worth $6bn per year\*. On the other hand, workshops are a big player for us and as of 2013 more than 15,000\* Mechanic shops have been operating in Australia. This means tools are needed to fulfil the market needs which are really high. Here is a hint: more than 20M vehicles are operating in the country, almost one vehicle per person! Crazy isn’t it?

Getting the right tools is a complicated task and then finding them is a real challenge. Ozwide Tools is there to solve these issues for the mechanics and make their life easier.

Basically, the project is a stock control forecaster which is widely applicable to other industries such as, retail, automotive dealerships, spare parts etc.

# Stakeholders

The main stakeholder is Ozwide Tools owner David McLeod.

# Business question

Is it possible as Ozwide tools to predict product sales for a certain amount of time in the future?

Why is this an important question?

This question has to be asked to reach the main goal which is how many products Ozwide Tools needs to purchase and hold in stock to fulfil the market needs of spending the least amount of money on freight for the next 6 months in time.

Getting a good forecasting model to work properly will save money and improve the profit per product.

# Data question

Can a model based on the data acquired by the internal system and external data such as marketing campaign signals predict the product demand of the market to forecast purchases and reduce the cost of freight?

# Data

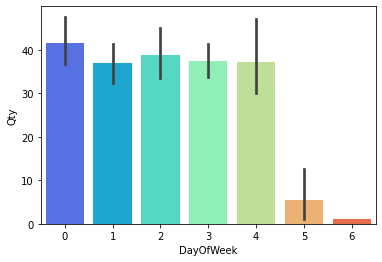
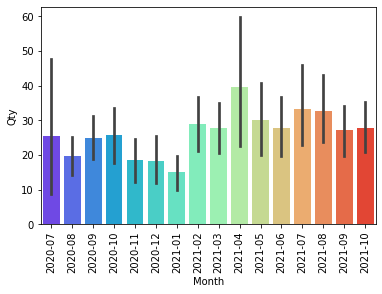
The data is gathered from the internal system Mpower. A report is generated by the system with the historical transactions per invoice.

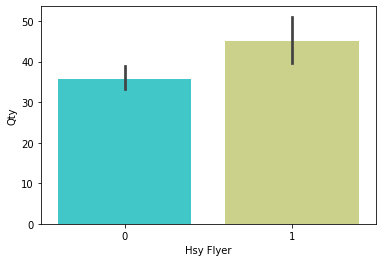
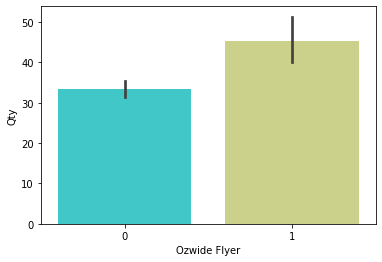
The dataset used includes 1550 records and 25 features divided in quantity, part number, supplier name, customer name, transaction type, description, timestamp and others.

The data is 100% reliable since iit is gathered directly from the internal system and it can be generated at any time which makes the model to be reproduced over the time.

# Data science process

## Data analysis

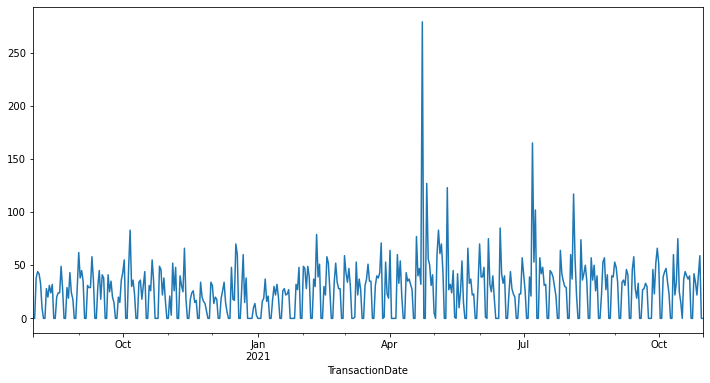
1. What data pipeline was to wrangle the raw data?
   1. Generating the data with the internal system on a CSV file
   2. Loading the data set into the environment
   3. Subsetting the data by the transaction types of interest which are sales and credits
   4. Resampling the data on a daily basis summing up the quantities sold and/or credited by day to get total quantities.
2. What are the highlights of the Exploratory Data Analysis (EDA)?
   1. During the EDA we found that the daily distribution of quantities during weekdays is showing an even distribution with a slight difference on Mondays and Wednesdays which are the days where more sales are happening we will take this into consideration down on the track
   2. We made the same process for monthly distribution finding that at the end of the financial year period there is an increment in sales
   3. Due to domain knowledge, we applied an external function that describes marketing campaigns applied on certain dates, we found out that these campaigns affect the behaviour of quantities sold with a 10% increase in sales as shown below



1. This pipeline is a reusable one based on the dataset implemented since the data structure doesn’t change on time, the files generated from the system keep the same feature and records structure.

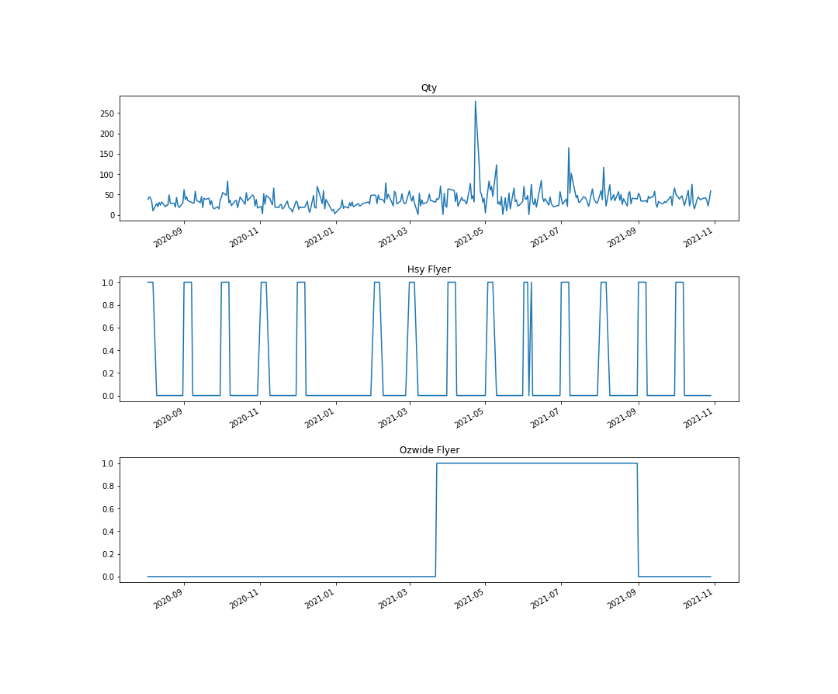
## Modelling

During the modelling process, we selected just the quantities out of the dataset as the main feature afterwards we subset the data by transaction type sale and credit we summed up the quantities and we ended up with a univariate dataset.



This graph describes the Quantities (y-axis) sold in the period of time of the x-axis.

During the EDA process as described above, we found out that the day of the week, day of the year and month can add insight into the model, describing seasonalities based on the time of the year. On the other hand based on experience we know that the times a marketing campaign is released the sales will boost, that is why it generates an input signal with the start and end date of flyers coming from Ozwide Tools and distributors.

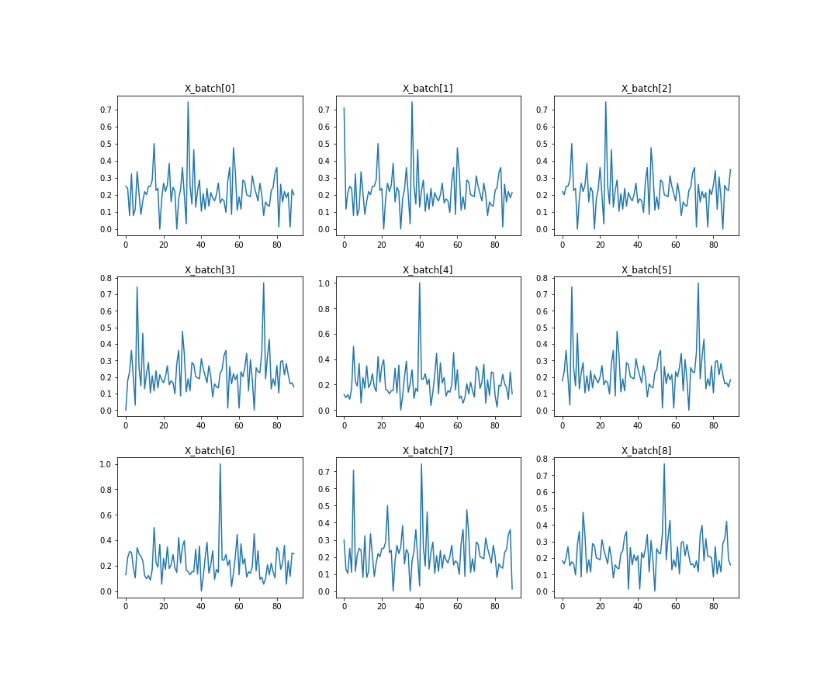


As pictured above, we can notice that there are some pikes coming after flyers. We assumed these signals are going to be a good input to catch those trends we want the model to understand.

At the end the features were selected by domain knowledge however, more inputs can be assessed and machine learning models such as Random Forest can be applied to get the best feature combination for the model.

After feature selection, to make the model work we had to turn it into a supervised one, in the case of time series the way to do this is shifting the results we are looking for the desired time steps in the future, this way the model will assess the training data based on seen data in the recorded future.

Finally, since we were applying neural models such as RNN and CNN scaling and batching the data was needed so the model understood the data.



Batched data scaled from 0 to 1

The models applied are listed below

1. Recurrent Neural Network (GRU)



Processing time 1 min 17 sec

1. Recurrent Neural Network (LSTM)



Processing time 45 sec

1. 1D Convolutional Neural Network



Processing time 40 sec

## Summary

| **Model** | **TRAIN RMSE** | **TEST RMSE** |
| --- | --- | --- |
| RNN GRU | 1.76 | 34.23 |
| RNN LSTM | 2.041 | 18.82 |
| 1D CONV | 6.99 | 6.41 |

The faster model was RNN LSTM and was the selected.

## Outcomes

* More year data is needed to build a more robust model.
* RNN and CNN are highly demanding models to train when it comes to computational power.
* Since this model is applied to the supply chain there is a broad field to implement it in different industries.
* RNN and CNN are a dynamic approach to time series forecasting since external factors can be taken into account when modeling them.

## Next Steps and Implementation

* Implement a RandomForest model for feature selection to improve the model.
* Extend the analysis to not just gross qty but to qty per supplier and product so we can forecast each product demand.
* Deploy the new model on a web service to calculate stock purchase based on product sales, marketing promotions, website click rate etc.

# Data answer

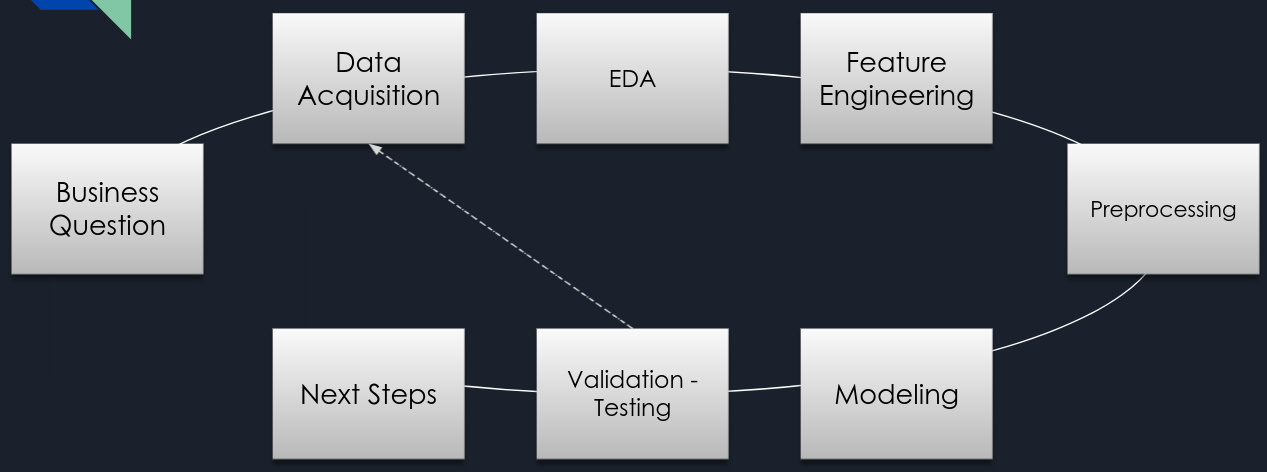
Based on the database the quantity forecasting is possible using all the three models applied, the results were promising and any of the models can be implemented, we have decided to use the RNN LSTM since it is the faster model and less demanding.

# Business answer

It is possible to forecast the stock based on the known data and marketing signals, since these neural network models are dynamic more inputs can be used to better describe the trends and make the neural network to understand better.

# End-to-end solution

Gathering the daily transactions report from the internal system Mpower and applying the pipeline



We can get a reliable forecasting system to calculate and program the stock purchases.

# References

**Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition**

<https://www.oreilly.com/library/view/hands-on-machine-learning/9781492032632/>

<https://github.com/ageron/handson-ml2/blob/master/15_processing_sequences_using_rnns_and_cnns.ipynb>

<https://github.com/Hvass-Labs/TensorFlow-Tutorials/blob/master/23_Time-Series-Prediction.ipynb>