A study of lap-times at the Nürburgring

# Introduction

As a car enthusiast, I am often looking for the latest trends in motorsport engineering. In automotive industry, performance and prestige is measured by lap time. The dataset and problem that I am targeting to solve is a model to predict lap times at the famous Nordschleife – Nürburgring. By creating this model, I hope to discern what are the greatest influencers to lap time and how car manufacturers can better improve their cars.

# Literature Review

Several searches and queries were made to try and find relevant research concerning automotive performance; however, ALL searches have yielded zero results. The searches were done through JSTOR, Google Scholar, and through other search engines. The only sources of information appear to be informal and not a result of academia.

The sources that I have consulted are automotive journaling magazines as well as interviews from various racing drivers. While no scientific research (concerning this topic) has been published or made publicly available, there is an overwhelming consensus from these field “experts”.

There are five major external influencers to lap time:

1. Tyre – How much contact surface area is available, and how does it perform when hot?
2. Engine – Horsepower, torque, and other engine characteristics.
3. Aerodynamics – How much downforce is the car able to generate at \_\_\_ speed?
4. Weight – How much does the automobile weigh?
5. Brakes – How much stopping force is available? How does performance suffer when hot?

All five of these metrics are scientifically measurable. However, the data is simply not available for brake performance, aerodynamic pressure, or tyre performance. It cannot be understated that racing is a form of sport; any scientific measurement applied to sport must be taken with a grain of salt.

There are internal influencers as well:

1. Driver confidence – How familiar is the driver with the car/instrument?
2. Driver skill – The ability to place the car at optimal entry and exit points (of a turn).
3. Track familiarity – How familiar is the driver with the track?

Regarding these three factors, we can disregard the third statement as all laps were done at a single track. Driver confidence and driver skill are also negated in this instance as manufacturers would have their own test drivers (peak confidence, skill, and familiarity), race the cars.

# Dataset

Give the description of the dataset that you are using along with the individual attributes you will or will not use in your analysis. Also mention the source of the dataset (where did you get it from). In case the data is curated and created by you please explain the details. Descriptive statistics of the attributes and datasets can also be provided here.

The original dataset is from Kaggle (<https://www.kaggle.com/scottdchris/nurburgring100/kernels>), and featured five variables (Position, Time, Year, Make, Model). The data featured 100 entries and was likely scraped either from a manufacturer website, or a third party site such as <https://fastestlaps.com/tracks/nordschleife>.

I have removed several entries from the dataset; they will not be considered for these reasons:

* Small manufacturers
* Specialty manufacturers
* Not road/street legal
* Multiples (unless substantial differences)
* Modified Vehicles

I have also added 13 measurable and accurate variables that will hopefully allow us to investigate this problem even further. In total, there are now 81 entries and 18 variables, and zero missing data points.

In the future, I may add other features to this dataset such as brake material, tyre, weather, and maximum lateral force.

There have been some preliminary observations of the data set:

* Most cars are of the Coupe body type, the others being Sedan and Hatchback
* Lighter cars appear to have greater torque, manual transmissions and better laptimes
* AWD systems appear to perform quicker than their RWD and FWD counterparts
* Hybrids are featured as both the quickest and slowest but also heaviest cars
* As expected, Horsepower and Torque are heavily correlated, they also correlate with superior laptimes
* Position and time appear to be redundant, but because the spread between position is not equal to time, both may be considered
* KG/HP is simply a function of Weight divided by Horsepower

# Approach

## Step 1: Find Dataset

A search was done on Google and Kaggle to find the relevant dataset as well as any other assisting data.

## Step 2: Edit Dataset

The dataset was then edited to find supplemental information as well as remove unnecessary information. There were no missing values or incorrect values to report.

## Step 3: Preliminary/Exploratory Analysis

A brief look at the complete dataset will be undertaken. Any explicit and noticeable trends will be noted as well as a basic analysis of correlation, outliers, and basic summary statistics.

## Step 4: Building the Model

Multiple models and methods will be done at this stage.

1. A basic multiple linear regression (performed step-wise for dimensionality reduction).
2. Classification/clustering, the cars will be categorized into different clusters.
3. Classification/clustering, the cars will be categorized into different clusters. Then we will create a variable to represent the different classes/clusters. We can then apply linear regression; having accounted for cars of similar classes/archetypes.
4. We will apply this model to both position and time. A total of 6 models.
5. We may also have a recommender system, where based from the lap time desired (and some other characteristics), we can recommend specific metrics to meet that goal.

## Step 5: Testing and Tuning the Model

These models would then be evaluated through cross validation/fold methodology. I may also scrape additional cars to test the model. Any variables that are intuitively significant may be manually handicapped if the model is ineffective in representing such variables.

## Step 6: Report and Visualization

At this stage, we will beautifully visualize our findings using Tableau. The report will be a concise business style summary that highlights all the key findings and numbers.