[**Vignesh**](https://in.linkedin.com/in/vignesh-jayanth-7b481440) **Jayanth (Professional Projects)**

**LS61RJ, Leeds, United Kingdom**

**Email**: [vigneshjayanth00@gmail.com](mailto:vigneshjayanth00@gmail.com)

**LinkedIn**: [Profile](https://in.linkedin.com/in/vignesh-jayanth-7b481440) **Github**: [Profile](https://github.com/vigneshjayanth00/) **Tableau**: [Profile](https://public.tableau.com/profile/vignesh.jayanth#!/)

**Phone**: **07871786016**

1. **Total Performance Data: Creation of Multivariate Model**

**Brief:** Project exploring the relationships between the variables associated to a horse race were produced to help demonstrate the value of the sectional times and stride analysis. The project included clear graphs/tables showing how stride length and/or cadence changes with race/horse details, and associated examples from recent history where a specific racehorse had exhibited “tell-tale signs” that it was about to improve for some change in distance or other variable.

**Tech Stack:** Python- TensorFlow, Decision-Trees, MATLAB

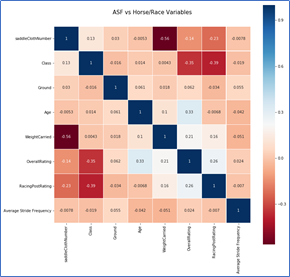
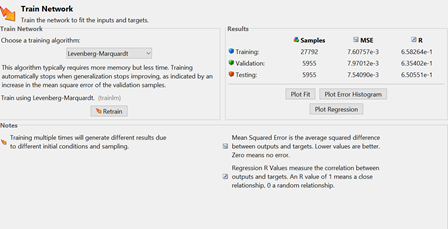
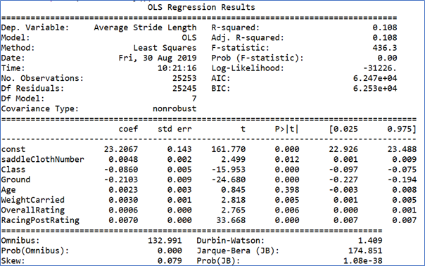
**Time Length:** Project time: 3 weeks, Data Size: All races from (2016-2019)- Flat races

**Process:**

* Data cleansing (Stride Frequency and Stride Length of horses for races)
* Application of different models on Python and Matlab

**Github:** <https://github.com/vigneshjayanth00/TPD>

**Snapshots:**

1. **Total Performance Data: Automation of Data Retrieval**

**Brief:** To extract back history and make it downloadable for B2C clients through the website for sales of back history data to armchair punters who frequently approach for the data. Using the back history to work on charts and tables to support a weekly blog to call out interesting speed/stride data

**Tech Stack:** Python- TensorFlow, Decision-Trees, Tableau (Visualization)

**Time Length:** Project time: 3 weeks, Data Size: All races from (2016-2019)- Flat races

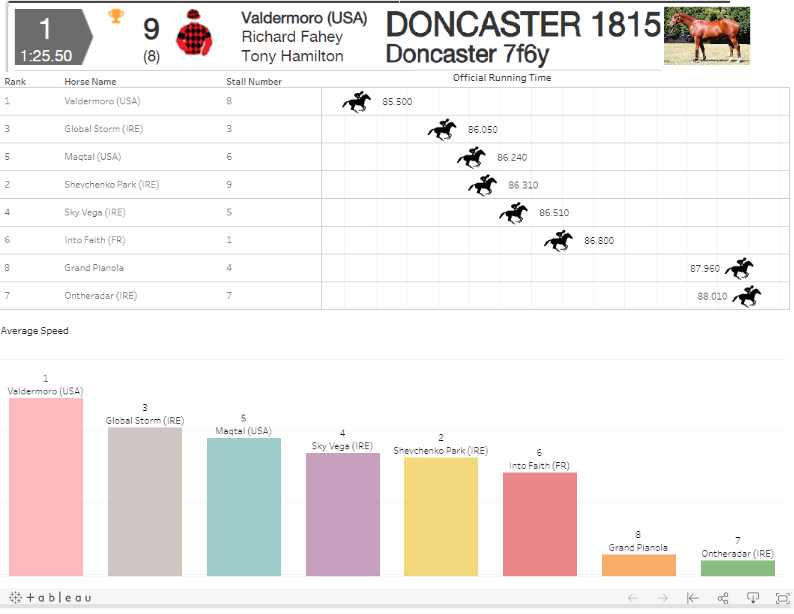
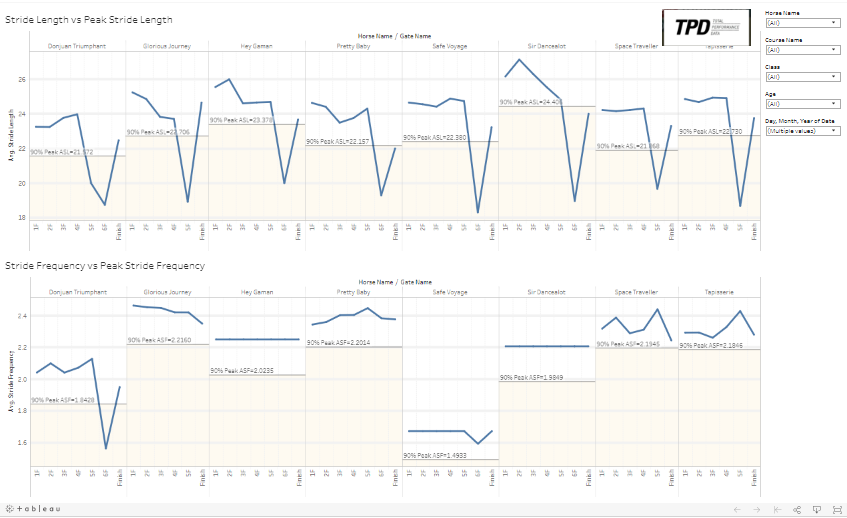
**Process:**

* Creation of efficient package to deploy for automation (various attributes in relation to horse races)
* Application of different models on Python to create charts on Tableau

**Github:** <https://github.com/vigneshjayanth00/TPD>

**Tableau:** <https://public.tableau.com/profile/vignesh.jayanth#!/>

**Snapshots:**

1. **Ssentif Analytics: How do physical assets influence crime in Leeds City?**

**Brief:** To investigate whether introducing physical assets (leisure centres, libraries, parks, bus routes) can decrease local antisocial behaviour and crime rates. Report on whether there is evidence for this effect. If there is, which physical asset is the most effective at decreasing antisocial behaviour and crime?

**Tech Stack:** MATLAB, Stats, Time-Series Analysis, Tableau (Visualization)

**Time Length:** Project time: 4 weeks

Data Size Limitations:

The area of library and leisure center overlap

The sample size is not large enough

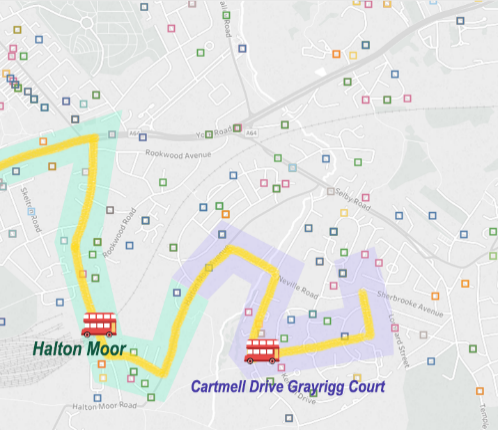
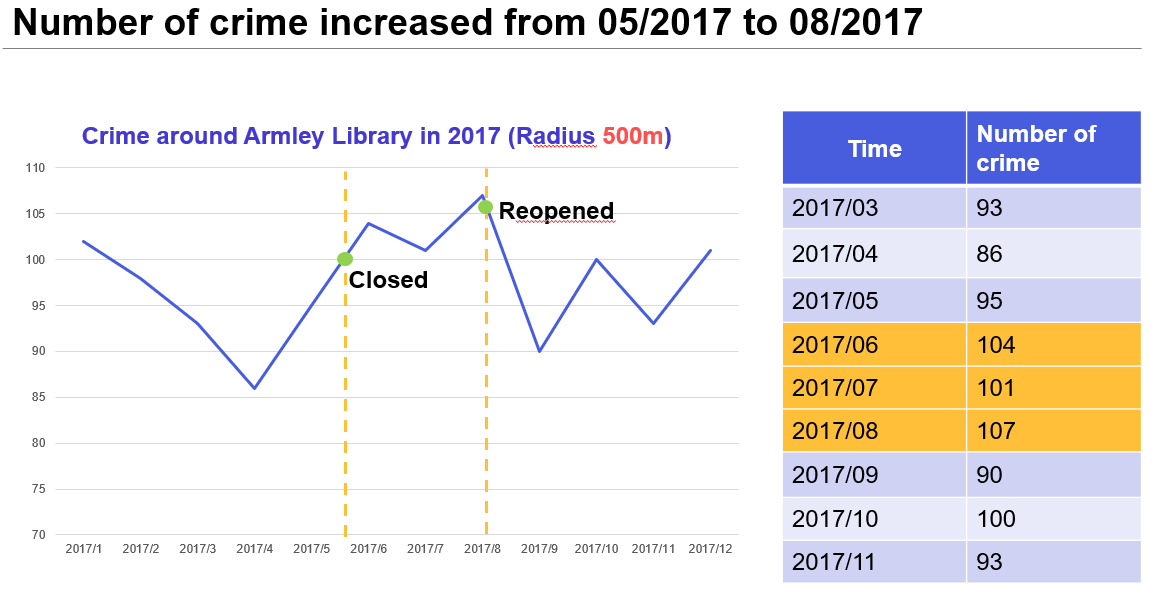
**Process:**

* Analyse physical assets namely library, bus stops and bus routes affect crime individually
* ANOVA test to verify trend in time series data for library
* Independent Samples two tailed test done on two samples for different bus routes

**Conclusions**

* **Library:** The existence of library might decrease the crime nearby
* **Leisure Centre:** The place closer to leisure center probably has higher crime density
* **Bus Stop:** Keep enough and healthy bus route operation may improve social safety

**Snapshots:**

**** 

1. **HSBC Analytics: Evaluating impact of incorrect customer balance statements due to technical glitch in mainframe systems**

**Brief:** The project aimed to identify those customers affected by a technical glitch in the mainframe systems. Incorrect balances (End-of-month) instead of (End-of-cycle) balances were sent to all customers for a period of 6 months due to a mainframe system error (The system picked up balances at the end of month due to a service management change at the time).

**Process:** An extensive analysis was carried out to reduce a potential loss of $20,000 for the business by identifying target customers affected by a technical glitch in the mainframe systems.

**Risk:** Potential risks avoided were disrepute to the organization, sending messages to all customers instead of only those affected (avoiding customer insecurity) and cost-to-company

The sample size included:

1. Flagging customers based on descending scale of being affected (Highly-Medium-Low)
2. Customers most affected by this were revolvers (revolving credit), identify this sample size to send personal messages (ideally those who transacted in that period avoiding dormant account)
3. The analysis was done by sub-setting data based on a negation technique to retain the final set of customers.

**Recognition: Service Excellence Champion 2016 Q1**



1. **Performance prediction for player recruitment in Premier League and Championship:**

**Brief:** The project aimed to identify players as suitable replacements for a premier league club from a database of players from the championship and the premier league.

**Data Inefficiencies:** Data had missing values for personal and player attributes and the sample size was two seasons for both (Championship and Premier League)

**Challenges posed:**

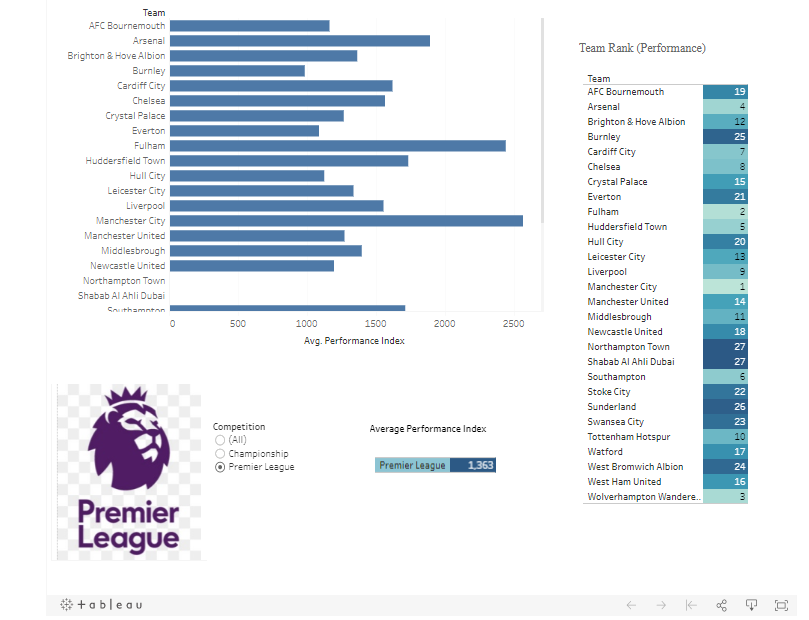
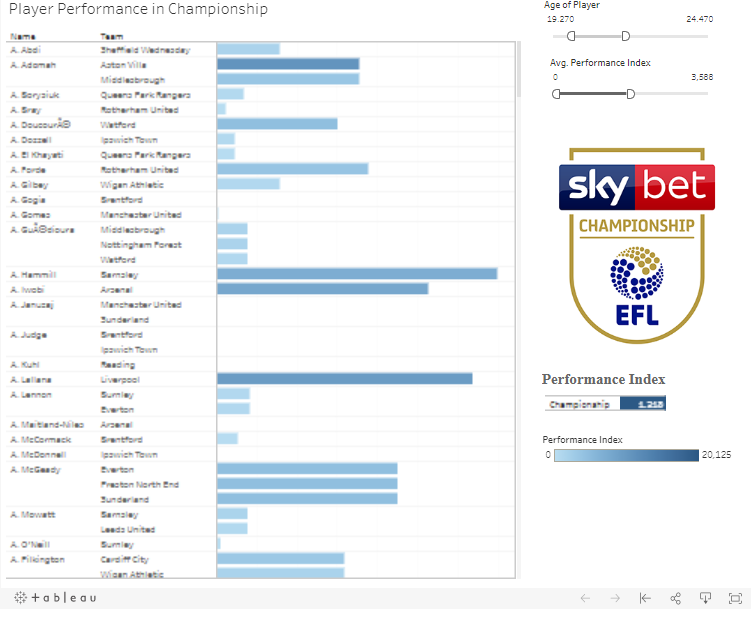
The players needed to be young, to maintain the team's age profile, At least two were to be English or homegrown, in order to meet the relevant squad restrictions and at least one needed to be from a Championship team, due to budget constraints

**Methodology**: A variable Performance Index was created using the Principal Component Analysis method (95% of features described) using performance metrics and categorical variables such as position, age, team, competition. The KNN classifier was used to predict missing data (Age, Nationality) with an accuracy of 70% (Predicted model of variables run on the training set). The Prediction technique used was K-neighbors Classifier. Training Set = 315 (Non-null), Test set=4585(Null). Accuracy~69% of 3 n\_neighbors chosen between the predicted Ages and ages in the training set. Data was subset after imputation of missing values and modelled based on the position of the outgoing player. Kernel Principal Component Analysis (PCA) method was done on all the performance metrics (excluding position-specific metrics) – 5 Features that contributed to 95% of the variance ratio for both midfielders and defenders. Average weighted Magnitude was combined with each performance metric and averaged to get the Average Performance Index.

**Conclusion & Limitations:** The belief is that, with a larger sample size and completeness of data, usage of PCA to impute values would have been a better option. K-NN adopted was a good method to identify and cluster players based on attributes.

**Link:** <https://public.tableau.com/profile/vignesh.jayanth#!/vizhome/PlayerPerformanceinChampionshipMidfielder/ChampionshipbyAge>

**Snapshots:**



GitHub: <https://github.com/vigneshjayanth00/Football>

**University Projects**

1. **Data warehouse Implementation: University Case Study**

**Brief:** Project implementing Star Schema, ETL processes, evaluating effectiveness of various Data warehouse architectures

**Tech Stack:** PL**-**SQL, Oracle APEX, QSEE SuperLite, OLAP- Cognos PowerPlay

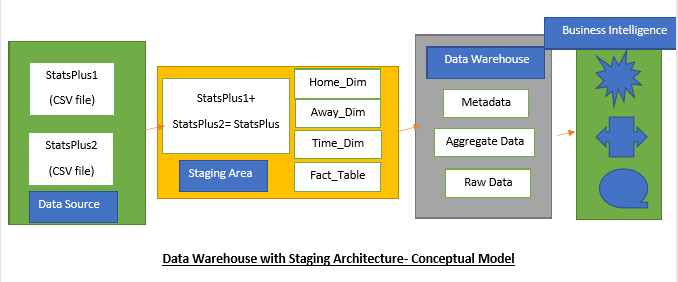
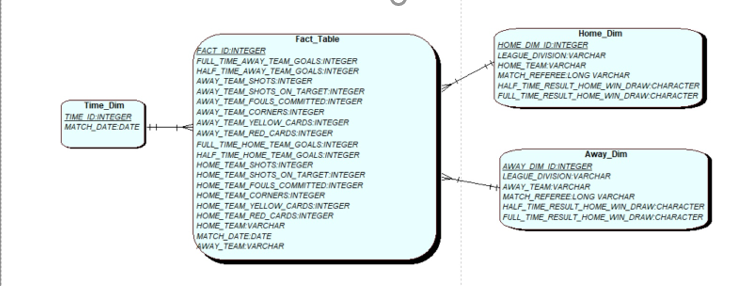
**Time Length:** Project time: 5 weeks, Data Size: 1400-2000 rows

**Process:**

The original datasets in the study used were smaller datasets that contained over 1400 and 2000 rows respectively. These were used to feed the dimension and fact tables from the Database source. The Data was fed into the APEX environment using the upload CSV file function under the utilities tab. A star schema was created using the QSEE SuperLite tool. The star schema consisted of one fact table, two-dimension tables and one-time dimension table. The Star schema was not in its normalized form. The data was populated in the fact and dimension tables using the upload data functionality under the utilities tab from the source data (Original Dataset). The functionality allows the user to choose the columns required to be populated (Eg: Facts for facts table). OLAP cubes were created using pivot tables on Cognos PowerPlay. After the scope of the project is defined and the requirements analysis is done, the selection of methodology is a requisite. By referring to the pros and cons of the Kimball/Inmon approaches mentioned in **(El Sappagh et al,2011, (Moscoso-Zea et al,2016),** it gives a good understanding of the most suited approach. Data Warehouse Architecture: Based on the survey study by **(Ariyachandra & Watson, 2008)**, Inmon and Kimball adopted either the hub or bus and spoke architecture. This case study adopted the Bus Architecture

**Grade:** Distinction

**Snapshot:**



1. **Database Design: University Case Study**

**Brief:** Project implementing Data Security in DB Design, Evaluation of SQL vs NoSQL Database Approaches: Database Implementation

**Tech Stack:** PL**-**SQL, Oracle APEX- Database Security, QSEE SuperLite- Star Schema

**Time Length:** Project time: 5 weeks

**Findings:**

**Database Security Controls:** Based on the report of data security, database security controls were put in place on the application suite. Primarily four areas of security measures were incorporated:

* Authentication Schemes
* Authorization Schemes
* Session State Protection
* Build Options

**Differences between SQL and NoSQL Database Approaches**

|  |  |
| --- | --- |
| SQL Engine vs NoSQL Engine Similarity | SQL Engine vs NoSQL Engine Difference |
| Examples of engines such as MongoDB and Oracle are database management systems that are cross-related to multiple programming languages | Data in NoSQL engines stored using dynamic table schemas as compared to fixed schemas in SQL engines |
| Examples of engines such as MongoDB and Oracle support multiple user databases | NoSQL engines are based on ‘document-oriented schema-less’ database model as compared to a relational model in SQL Engines |

Some key differences between Mongo DB and Oracle DB was suggested based on the study done by **(Mayur M Patil et al, 2017):**

|  |  |
| --- | --- |
| **Mongo-DB** | **Oracle- DB** |
| Schema-Less database model | Relational Database model |
| Query Language consists of API calls | Query language is SQL |
| Value size is 16MB | Value size is 4KB |
| Accepts large amounts of data | Performance is slow for large amounts of data |
| Uses functions for operations such as adding, deleting and updating | Uses SQL language to insert, select, update |
| Uses ‘callback’ functions for advanced queries | Uses PLSQL language for advanced querying |
| It is open source product | It is paid licensed product |

**Grade:** 2:1

**Snapshot**