# City Transport Planning

|  |  |
| --- | --- |
| Project description | A large city wants to pilot a new scheme to use data science to look at transport use in their area. Commuters in the city are now using Oyster-like cards to tap-in and out on their journeys. This data can now be collected in real-time and used for analysis.  The city wants to use this data to:   * Understand the daily and weekly transport demand cycle * Gain an understanding if there are trends in transport usage * Plan for further transport needs * Gain an understanding of what would be required to scale-up this pilot to a full-size platform looking at several million journeys per day   No personal or sensitive data will be captured.  The city is keen for non-technical users (i.e. staff and public) to be able to interact with the data and to understand its meaning.  The city would like to keep costs as low as possible, but wants to build a robust and expansible platform.  The implementation of the system must be auditable and include structured reporting to management. |
| Challenge | Propose and describe a solution that would enable a system to be implemented to store and analyse the city’s travel data.  Select the staff and technology needed to support the solution.  Describe how the system would be expansible over the coming years. |

## Online Retail Fraud Prevention

|  |  |
| --- | --- |
| Project description | A nationally-prominent online retailer has wants to reduce the number of fraudulent transactions that are serviced by their applications. They also want to reduce the number of manual checks that are performed. The retailer believes that a good understanding of potentially fraudulent transactions can be gained through the analysis of existing data.  The retailer’s existing fraud team understand the factors that increase the probability of a transaction being fraudulent (e.g. Delivery to non-residential property).  The retailer’s senior management want you to implement a system that examines transactions as they take place, flags potentially fraudulent transactions for manual checks. The management want the system to learn from experience so that automatic checks can replace manual checks. The system must not delay transactions unnecessarily and must keep false positives to an absolute minimum.  This project will require the use of personally identifiable data.    The decisions made by the system must be explicable.  The project must have a working version in place quickly. The working version need not be the final version, but must indicate how an automated solution would work. |
| Challenge | Propose and describe a solution that would enable fraud checks to be made on the retailer’s transactions. Explain how the data would be analysed, how existing knowledge would be incorporated, how the system would improve its performance.  Explain how sensitive data would be managed and how the system would be implemented.  Select the staff and technology needed to support the solution. |

## Customer ‘journey’ analytics

|  |  |
| --- | --- |
| Project description | A company that provides online video games has launched a new game and wants to understand its customers’ level of engagement and where they are on their customer journey.  The new game is capable of generating a massive amount of data describing players’ behaviour. The company’s existing system for storing and analysing behavioural data is being quickly overwhelmed – existing databases are becoming unusable and most data is not being analysed.  The company wants to detect where customers are losing interest in this game so that they can intervene. The company wants to gain an understanding of customers’ interests so that they can make recommendations of other products to enhance engagement.  The company wants to implement an analysis system for the game that can be subsequently be applied to other products.  The company wants to understand what would be an individual customer’s total lifetime value to enable budgeting and to highlight where effort to retain customers should be targeted.  The meaning of the analysis must be made clear to non-technical staff.  The company does not have its own datacentre and all services are operated from an Amazon Web Services environment.  The project will entail the use of sensitive personal data. |
| Challenge | Propose and describe a solution that would enable game data to be analysed and for use behaviour to be understood.  Explain how recommendations could be made both for interventions to enhance engagement and to encourage users to play additional games.  Explain how lifetime value might be assessed.  Explain how sensitive data would be managed and how the system would be implemented.  Select the staff and technology needed to support the solution. |

## 

## Deep learning methods for classifying aerial photography

|  |  |
| --- | --- |
| Project description | An environmental agency is interested in learning about the changing characteristics of land usage in Britain. The agency has an extensive collection of aerial photographs and satellite images of Britain and wants to develop a system that can classify land usage and changes of interest.  The agency would like to develop a system that can initially make use of specialist knowledge and then learn through experience so that it can perform as well as a human specialist without intervention.  The agency holds many thousands of image files, each of which is of a large size.  The data is not considered sensitive.  The agency does not care whether the analysis is performed on-site or in a cloud environment.  The agency believes that system performance may be enhanced by blending data from multiple sources into a single analysis.  The agency prefers the use of open-source tools to control costs.  The system’s classification rationale doesn’t have to be explicable provided it is effective. |
| Challenge | Propose and describe a solution that would enable the contents of the images to be classified.  Explain how expert knowledge will be captured by the system.  Explain how the system will improve performance.  Explain where the computing will take place.  Select the staff and technology needed to support the solution. |