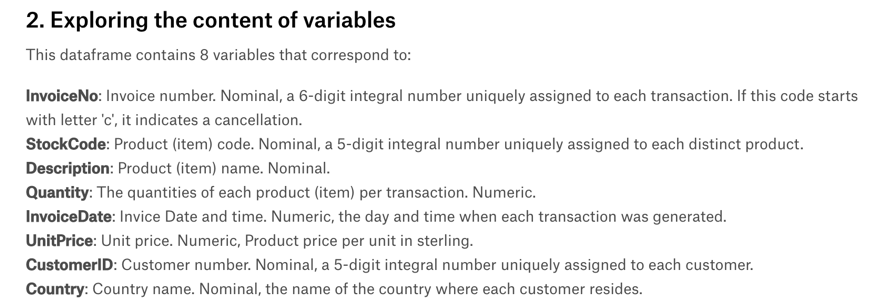
1. Informations sur le DataSet
   1. Préparation
      1. Quoi ? Entreprise dans le UK sur le Online Retail
      2. Nb d’enregistrement (length)
      3. Periode (head(date) et tail (date)
   2. Contrôle des données
      1. Combien de variables ?
      2. Quelles sont les variables et leur format ?



* + 1. Exploration
       1. NB de countries
       2. Cb de customers et transactions
       3. Cancelling orders
       4. Stockcode
          1. POST
          2. D
          3. C2
          4. M
          5. Bank Charges
          6. PADS
          7. Dots
  1. Data Clean
     1. Supprimer les données non nécéssaires ('POST', 'D', 'C2', 'M', 'BANK CHARGES', 'PADS', 'DOT')
     2. Supprimer les données dupliquées
     3. % de données supprimés: New Data length (dim(OnlineretailClean)-dim(UniqueOnelineretailClean))

1. Descriptive Statistics
   1. Data visualisations
      1. Number of unique invoices
      2. Number of unique products
      3. Number of unique customers
      4. Number of Country
      5. Summary of Quantity
   2. Some Graphs
      1. Invoices
         1. Per months
         2. Per weeks
         3. Per hours
      2. Turnover
         1. Per month
2. PCA
   1. Country dataset
      1. Pairs(countryData)
      2. Use PCA (with scale() and princomp() )
      3. Summary(countryData)
   2. Country without UK dataset
      1. Pairs ()
      2. Use PCA (with scale() and princomp() )
      3. Summary(*pcaCountryWithoutUK*)
   3. Product dataset
      1. Pairs ()
      2. Cor()
      3. Use PCA (with scale() and princomp() )
      4. Cov()
      5. Summary(*pcaProduct*)
      6. Plot(pcaProduct)
      7. Utilisation ade4
         1. inertia.dudi(pcaProductAde4,col.inertia = T)$col.abs
         2. score(pcaProductAde4, xax=1)
         3. score(pcaProductAde4, xax=2)
         4. s.corcircle(pcaProductAde4$co)
3. Clustering
4. Analyse des données importées
   1. Qui? Magasin au UK ...
   2. Quoi? On décrit les variables du dataset (length + fonction qui décrit les variables)
   3. Ou? UK
   4. Quand? On utilise la fonction head et tail pour montrer la période couverte
   5. Combien? Nombre de lignes dans le data set (fonction dim)
5. Nettoyage des données
   1. Données non pertinantes
      1. C, C2, D, POST, Price < 0, Quantity <0, Dupliquées, ...
   2. Nombre de lignes Nettoyées
      1. Nombre de lignes au total
      2. % par rapport au dataset original
6. Statistiques Descriptives
   1. Analyse Générale
      1. Nb de pays
      2. Nb de factures
      3. Nb de clients
   2. Analyse des ventes selon les pays
      1. Piechart avec UK en comparaison avec les ventes en dehors
      2. Piechart des ventes en dehors de UK
      3. Piechart avec % ventes hors UK
      4. Pays avec le plus grand nombre de retour
   3. Analyse des ventes
      1. Summary (summary(purchasesPerCountry[2]), summary(salesPerProduct[2]))
      2. Boxplot des ventes (2 boxplots : quantity et sales per products)
      3. Month Graph
      4. Day Graph
      5. Hour Graph
7. PCA
   1. DataSet par pays (pairs())
      1. Plot du pca pays => non pertinent
      2. Plot du pca pays (sans UK) => non pertinent
   2. DataSet par produit (pairs())
      1. Matrice des corrélations
      2. Sommaire des composants (summary(pcaProduct))
      3. Plot du pca (plot(pcaProduct))
      4. Poids des variables originales dans les composantes (Loadings(pcaProduct))
      5. Lien entre les composantes et les variables (score(pcaProductAde4, xax=1) et score(pcaProductAde4, xax=2))
      6. Cercle de corrélation (s.corcircle(pcaProductAde4$co))
8. Clustering