

# Online Hackathon On Data Driven Innovation For Citizen Grievance Redressal

- Prateek Khanna

# DataSets

- Public Grievance details in CPGRAMS along with feedback details
- Public Grievance movement details across the organization in CPGRAMS
- Monthly Department-wise public grievance receipts and disposals from January 2016 to October 2019
- Department-wise receipts, disposal and pendency of Public Grievance detailed statistics from 01.01.2016 to 01.11.2019
- Details of registered users of CPGRAMS as on 24.10.2019
- CPGRAMS Nodal Public Grievance Officers Details as on 30th October 2019

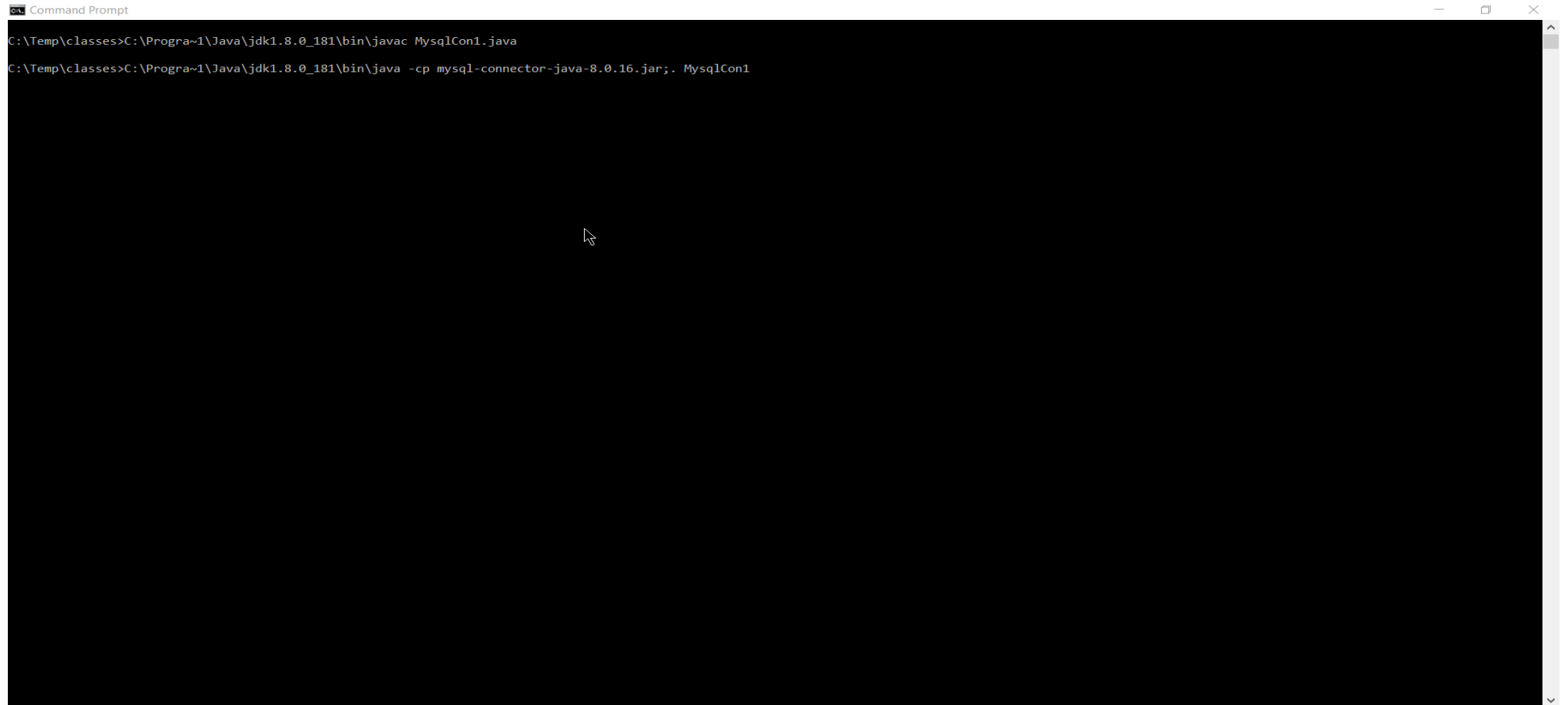
# Predictive Models

- The implementation falls under the category of Application Predictive Models.
- It makes use of the shared datasets and predicts future trends for the relevant metrics.
- The time series data is analysed Auto Regressive Integrated Moving Average (ARIMA) methodology.
- Data is populated in a MySQL database and java code is used for generating relevant csv files.
- These csv files are then used as input to a R program to generate future predictions.

# Setup

- Source code repository :  
<https://github.com/solfinder/predictivemodels>
- Deployment steps:  
<https://github.com/solfinder/predictivemodels/blob/master/README.md>
- Java Class : MySqlCon1
- R module : darpg.r
- Time Series Predictive Model : ARIMA
- Prediction data output format : csv files containing data for future months

# Predictive Models – Java Preprocessing































```
Command Prompt
C:\Temp\classes>C:\Program Files\Java\jdk1.8.0_181\bin\javac MysqlCon1.java
C:\Temp\classes>C:\Program Files\Java\jdk1.8.0_181\bin\java -cp mysql-connector-java-8.0.16.jar;. MysqlCon1
```





























# Predictive Models – Java Preprocessing

- Java Preprocessing Class MysqlCon1 connects to the database table containing raw data about grievance receipts and disposals and transforms it into a set of csv files organized by departments.
- Individual csv file is generated for each department for both receipts and disposal of grievances.
- These csv files are used as input for creating time series models for the respective departments.
- The csv files contain metrics from a start month to an end month.
- Prediction model computes the metrics for the subsequent months.

# Predictive Models – Java Preprocessing

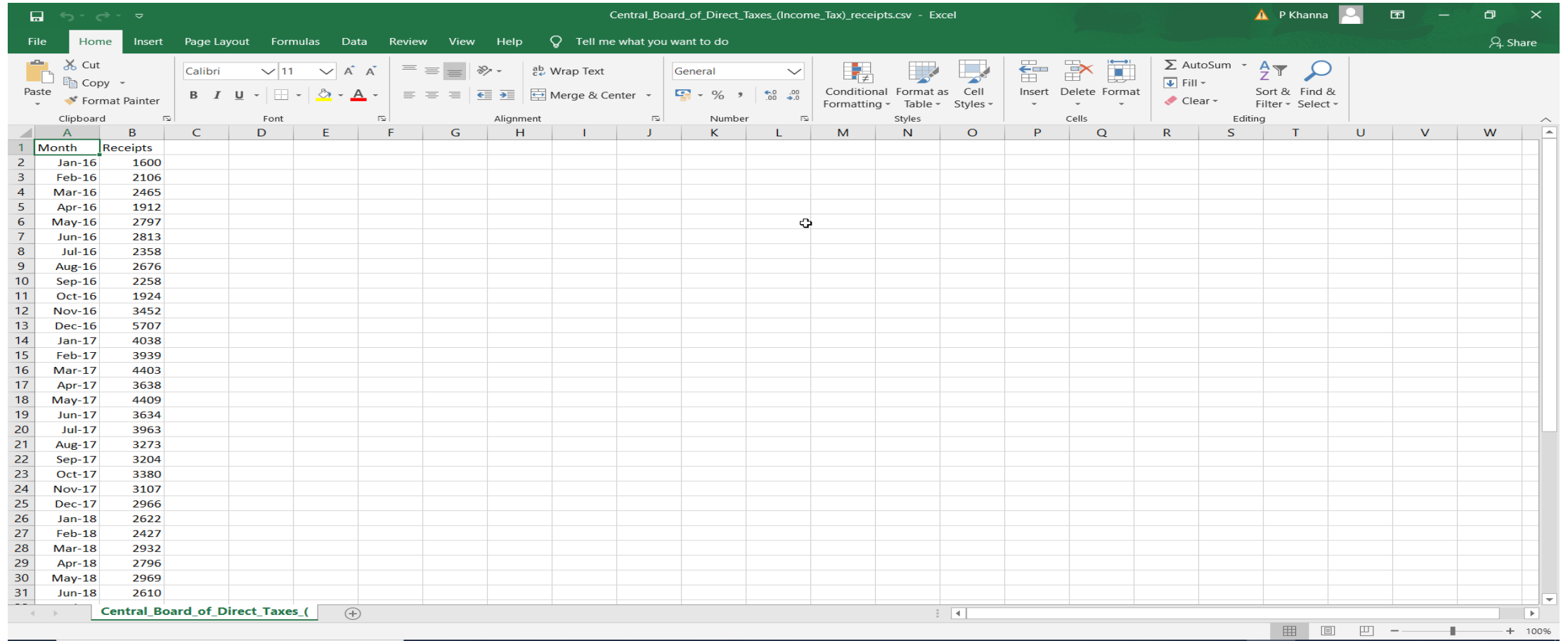
Name	Date modified	Type	Size
 Central_Board_of_Direct_Taxes_(Income_Tax)_receipts.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Central_Board_of_Indirect_Taxes_and_Customs_receipts.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Committee_on_Petitions_Rajya_Sabha_receipts.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department_of_Administrative_Reforms_and_PG_receipts.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department_of_Agriculture_Cooperation_and_Farmers_receipts.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department_of_Agriculture_Research_and_Education_receipts.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department_of_Animal_Husbandry_Dairying_receipts.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department_of_Atomic_Energy_receipts.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department_of_Bio_Technology_receipts.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department_of_Chemicals_and_Petrochemicals_receipts.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department_of_Commerce_receipts.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department_of_Consumer_Affairs_receipts.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department_of_Defence_Finance_receipts.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department_of_Defence_Production_receipts.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department_of_Defence_receipts.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department_of_Defence_Research_and_Development_receipts.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department_of_Economic_Affairs_ACC_Division_receipts.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department_of_Empowerment_of_Persons_with_Disabili_receipts.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department_of_Ex_Servicemen_Welfare_receipts.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department_of_Expenditure_receipts.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department_of_Fertilizers_receipts.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department_of_Financial_Services_(Banking_Division_receipts.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department_of_Financial_Services_(Insurance_Divisi_receipts.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department_of_Financial_Services_(Pension_Reforms)_receipts.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department_of_Fisheries_receipts.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department_of_Food_and_Public_Distribution_receipts.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department_of_Health_& Family_Welfare_receipts.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department of Health Research receipts.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB

# Predictive Models – Java Preprocessing

Name	Date modified	Type	Size
 Central_Board_of_Direct_Taxes_(Income_Tax)_disposal.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Central_Board_of_Indirect_Taxes_and_Customs_disposal.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Committee_on_Petitions_Rajya_Sabha_disposal.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department_of_Administrative_Reforms_and_PG_disposal.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department_of_Agriculture_Cooperation_and_Farmers_disposal.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department_of_Agriculture_Research_and_Education_disposal.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department_of_Animal_Husbandry,_Dairying_disposal.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department_of_Atomic_Energy_disposal.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department_of_Bio_Technology_disposal.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department_of_Chemicals_and_Petrochemicals_disposal.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department_of_Commerce_disposal.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department_of_Consumer_Affairs_disposal.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department_of_Defence_disposal.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department_of_Defence_Finance_disposal.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department_of_Defence_Production_disposal.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department_of_Defence_Research_and_Development_disposal.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department_of_Economic_Affairs_ACC_Division_disposal.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department_of_Empowerment_of_Persons_with_Disabili_disposal.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department_of_Ex_Servicemen_Welfare_disposal.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department_of_Expenditure_disposal.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department_of_Fertilizers_disposal.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department_of_Financial_Services_(Banking_Division_disposal.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department_of_Financial_Services_(Insurance_Divisi_disposal.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department_of_Financial_Services_(Pension_Reforms)_disposal.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department_of_Fisheries_disposal.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department_of_Food_and_Public_Distribution_disposal.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department_of_Health_& Family_Welfare_disposal.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB
 Department of Health Research disposal.csv	09-01-2020 20:28	Microsoft Excel Co...	1 KB



# Predictive Models – Java Preprocessing



Central\_Board\_of\_Direct\_Taxes\_(Income\_Tax)\_receipts.csv - Excel

Month	Receipts
Jan-16	1600
Feb-16	2106
Mar-16	2465
Apr-16	1912
May-16	2797
Jun-16	2813
Jul-16	2358
Aug-16	2676
Sep-16	2258
Oct-16	1924
Nov-16	3452
Dec-16	5707
Jan-17	4038
Feb-17	3939
Mar-17	4403
Apr-17	3638
May-17	4409
Jun-17	3634
Jul-17	3963
Aug-17	3273
Sep-17	3204
Oct-17	3380
Nov-17	3107
Dec-17	2966
Jan-18	2622
Feb-18	2427
Mar-18	2932
Apr-18	2796
May-18	2969
Jun-18	2610

# Predictive Models – R TimeSeries Prediction Model

```
RGui (32-bit) - [C:\Temp\darpg.r - R Editor]
File Edit Packages Windows Help

# R Module for timeseries forecasting of received and disposed grievances
# Developed by Prateek Khanna

install.packages("forecast")
install.packages("Metrics")
# loading packages
library(forecast)
library(Metrics)
library(ggplot2)

#Default forecast period is set to 12
#Update this value to generate forecasts for different periods
forecast_period=12

datal = read.csv("C:\\temp\\classes\\csv_receipts\\NITI_Aayog_receipts.csv")

nm <- list.files(path="C:\\temp\\classes\\csv_receipts")
for(filename in nm){
  # reading data from csv file
  data = read.csv(paste("C:\\temp\\classes\\csv_receipts\\",filename,sep=""))
  #ggplot(data, aes(x=Month, y=Receipts)) + geom_point()
  #ggsave(paste(paste("C:\\temp\\classes\\csv_receipts_graph\\",filename,sep=""),".png",sep=""),width = 25, height = 25)
  # training data sets
  train = data[1:nrow(data),]

  # removing "Month" column
  train$Month = NULL

  # training model
  model = auto.arima(train)

  # model summary
  sink(paste(paste("C:\\temp\\classes\\csv_receipts_predictionmodel\\",filename,sep=""),".txt",sep=""))
  print(summary(model))
  sink()

  # forecasting
  forecast = predict(model,forecast_period)

  write.table(forecast$pred, file = paste("C:\\temp\\classes\\csv_receipts_prediction\\",filename,sep=""), sep = ",", col.names=NA)

  #ggplot(data2, aes(x=Month, y=Receipts)) + geom_point()
  #ggsave(paste(paste("C:\\temp\\classes\\csv_receipts_prediction_graph\\",filename,sep=""),".png",sep=""))
}

nm <- list.files(path="C:\\temp\\classes\\csv_disposal")
for(filename in nm){
```

# Predictive Models – R TimeSeries Prediction Model

```
RGui (32-bit) - [C:\Temp\darpg.r - R Editor]
File Edit Packages Windows Help

# R Module for timeseries forecasting of received and disposed grievances
# Developed by Prateek Khanna

install.packages("forecast")
install.packages("Metrics")
# loading packages
library(forecast)
library(Metrics)
library(ggplot2)

#Default forecast period is set to 12
#Update this value to generate forecasts for different periods
forecast_period=12

data1 = read.csv("C:\\temp\\classes\\csv_receipts\\NITI_Aayog_receipts.csv")

nm <- list.files(path="C:\\temp\\classes\\csv_receipts")
for(filename in nm){
  # reading data from csv file
  data = read.csv(paste("C:\\temp\\classes\\csv_receipts\\",filename,sep=""))
  #ggplot(data, aes(x=Month, y=Receipts)) + geom_point()
  #ggsave(paste(paste("C:\\temp\\classes\\csv_receipts_graph\\",filename,sep=""),".png",sep=""),width = 25, height = 25)
  # training data sets
  train = data[1:nrow(data),]

  # removing "Month" column
  train$Month = NULL

  # training model
  model = auto.arima(train)

  # model summary
  sink(paste(paste("C:\\temp\\classes\\csv_receipts_predictionmodel\\",filename,sep=""),".txt",sep=""))
  print(summary(model))
  sink()

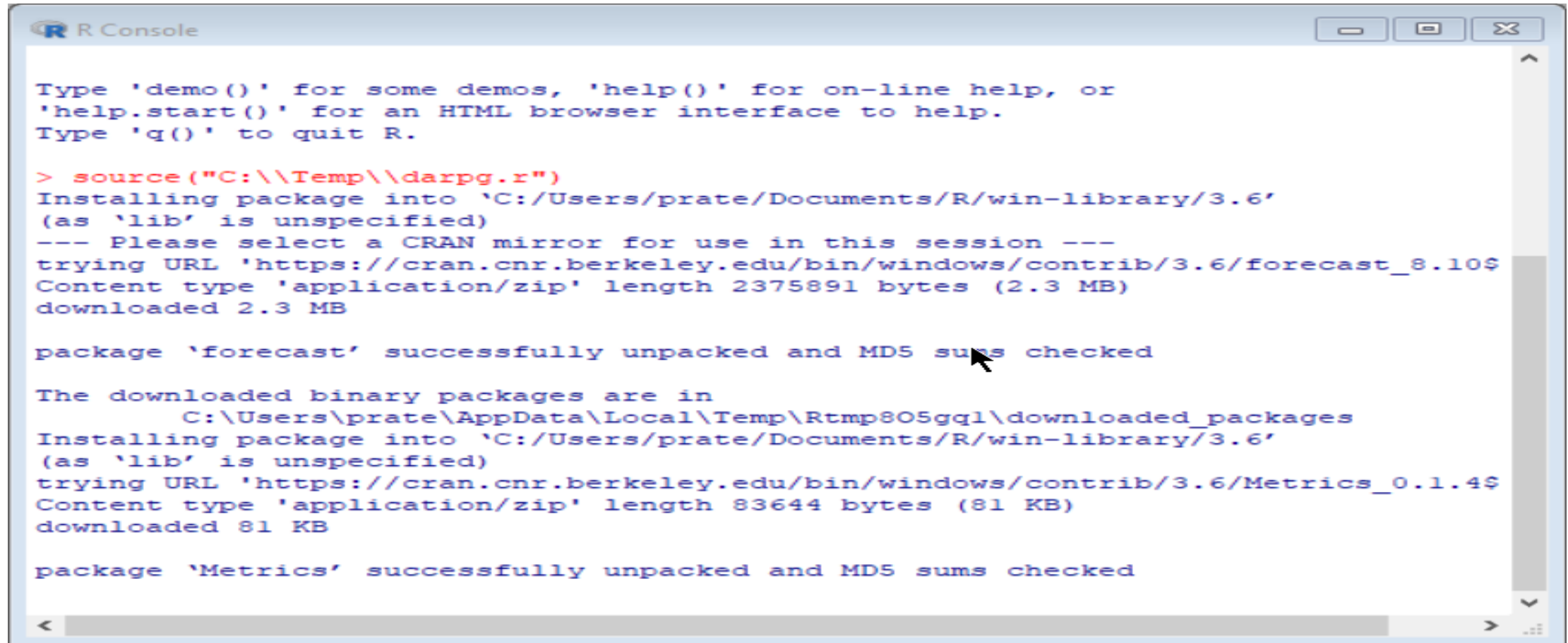
  # forecasting
  forecast = predict(model,forecast_period)

  write.table(forecast$pred, file = paste("C:\\temp\\classes\\csv_receipts_prediction\\",filename,sep=""), sep = ",", col.names=NA)

  #ggplot(data2, aes(x=Month, y=Receipts)) + geom_point()
  #ggsave(paste(paste("C:\\temp\\classes\\csv_receipts_prediction_graph\\",filename,sep=""),".png",sep=""))
}

nm <- list.files(path="C:\\temp\\classes\\csv_disposal")
for(filename in nm){
```

# Predictive Models – R TimeSeries Prediction Model



```
R Console

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

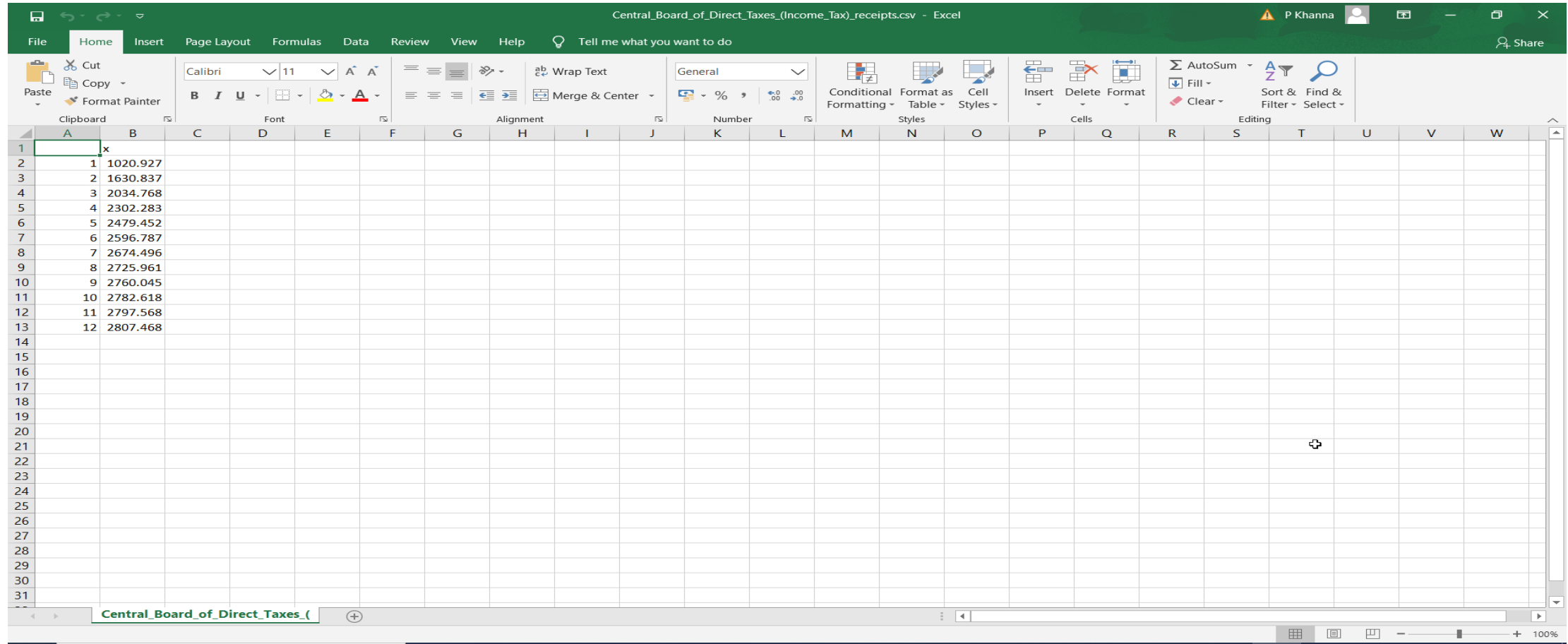
> source("C:\\Temp\\darpq.r")
Installing package into 'C:/Users/prate/Documents/R/win-library/3.6'
(as 'lib' is unspecified)
--- Please select a CRAN mirror for use in this session ---
trying URL 'https://cran.cnr.berkeley.edu/bin/windows/contrib/3.6/forecast_8.10$
Content type 'application/zip' length 2375891 bytes (2.3 MB)
downloaded 2.3 MB

package 'forecast' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
      C:\Users\prate\AppData\Local\Temp\Rtmp8O5gql\downloaded_packages
Installing package into 'C:/Users/prate/Documents/R/win-library/3.6'
(as 'lib' is unspecified)
trying URL 'https://cran.cnr.berkeley.edu/bin/windows/contrib/3.6/Metrics_0.1.4$
Content type 'application/zip' length 83644 bytes (81 KB)
downloaded 81 KB

package 'Metrics' successfully unpacked and MD5 sums checked
```

# Predictive Models – R TimeSeries Prediction Example



The screenshot displays the Microsoft Excel interface with the file "Central\_Board\_of\_Direct\_Taxes\_(Income\_Tax)\_receipts.csv" open. The ribbon is set to "Home", and the "Clipboard" group is active. The worksheet contains a time series dataset with two columns: an index (1-12) and tax receipt values. The data is as follows:

Index	Tax Receipts
1	1020.927
2	1630.837
3	2034.768
4	2302.283
5	2479.452
6	2596.787
7	2674.496
8	2725.961
9	2760.045
10	2782.618
11	2797.568
12	2807.468

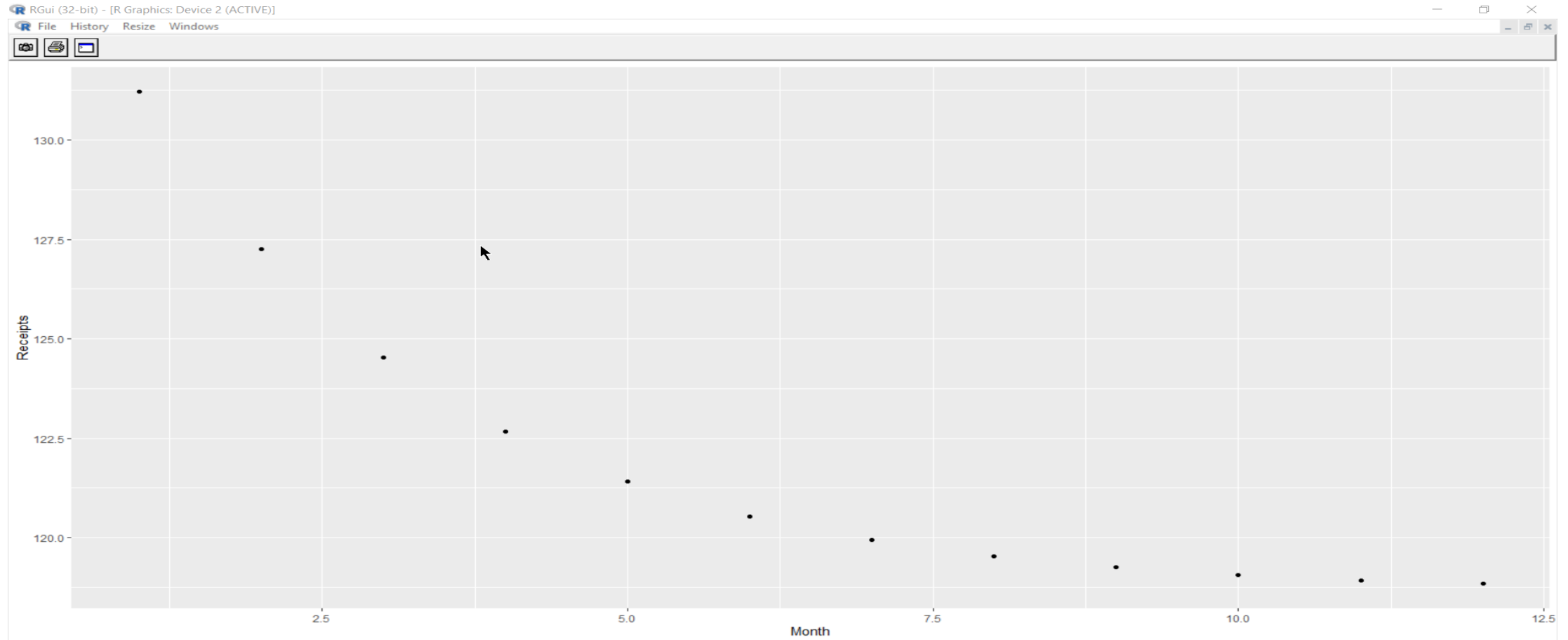
The Excel window title bar shows the file name and the user "P Khanna". The status bar at the bottom indicates the active sheet is "Central\_Board\_of\_Direct\_Taxes\_(". The zoom level is set to 100%.

# Predictive Models – R TimeSeries Prediction Example

The screenshot shows an Excel spreadsheet titled "Central\_Board\_of\_Direct\_Taxes\_(Income\_Tax)\_disposal.csv". The data is organized in two columns: a date column (A) and a numerical value column (B). The dates range from 1 to 12 of each month for the years 2016, 2017, and 2018. The numerical values represent tax amounts, starting at 1267.637 in January 2016 and increasing to 2774.238 in December 2018. The Excel interface includes the standard ribbon with tabs for File, Home, Insert, Page Layout, Formulas, Data, Review, View, and Help. The Home tab is active, showing options for Clipboard, Font, Alignment, Number, Styles, Cells, and Editing. The status bar at the bottom indicates the file is "Ready" and the zoom level is 100%.

	A	B
1		x
2	1	1267.637
3	2	1896.597
4	3	2263.688
5	4	2477.94
6	5	2602.988
7	6	2675.972
8	7	2718.569
9	8	2743.43
10	9	2757.941
11	10	2766.41
12	11	2771.353
13	12	2774.238
14		
15		
16		
17		
18		
19		
20		
21		
22		
23		
24		
25		
26		
27		
28		
29		
30		
31		

# Predictive Models – R TimeSeries Prediction Example



# Predictive Models – R TimeSeries Prediction Model Details

```
Department_of_Bio_Technology_receipts.csv.txt - Notepad
File Edit Format View Help
Series: train
ARIMA(0,0,0) with non-zero mean

Coefficients:
      mean
    20.9767
s.e.    1.6603

sigma^2 estimated as 121.4:  log likelihood=-163.68
AIC=331.36   AICc=331.66   BIC=334.88

Training set error measures:
              ME      RMSE      MAE      MPE      MAPE      MASE
Training set -1.817377e-15 10.88735  7.92861 -30.49891 52.00143 0.7909777
              ACF1
Training set 0.1981748

              ME      RMSE      MAE      MPE      MAPE      MASE
Training set -1.817377e-15 10.88735  7.92861 -30.49891 52.00143 0.7909777
              ACF1
Training set 0.1981748
```



# Predictive Models – R TimeSeries Prediction Model Details

```
Central_Board_of_Direct_Taxes_(Income_Tax)_disposal.csv.txt - Notepad
File Edit Format View Help
Series: train
ARIMA(1,0,0) with non-zero mean

Coefficients:
      ar1      mean
    0.5836 2778.2818
s.e.  0.1454  328.5908

sigma^2 estimated as 901085:  log likelihood=-388.09
AIC=782.19  AICc=782.74  BIC=787.74

Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 44.8403 928.8383 651.8077 -134.768 153.4398 0.9129238 -0.1436935
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 44.8403 928.8383 651.8077 -134.768 153.4398 0.9129238 -0.1436935
```

# Conclusion

- The application makes use of ARIMA modelling in R to predict future datapoints.
- Details of the actual model characteristics for each series are also captured and reported for reference.
- Predicted data can be used for:
  - Optimum resource utilization
  - Reallocation of resources is required
  - Investment strategy formulation
  - Contingency planning