**Business Context**

For this project, you need to step inside the shoes of a consultant and resolve the problems that your clients are facing. In this regard, you will be required to look at the problems being faced by two clients:

1. Golden Harvest, a startup dealing in gold lockets and charms
2. Human resources department of a services-based company

**Project Structure**

In tandem with the number of clients, the project is divided into two parts:

1. Part 1: Solving for Golden Harvest
2. Part 2: Helping human resources

You will be provided with the broad problem statements and datasets for the two parts in the next segment, followed by the Jupyter notebooks for them.

**Assessment Criteria**

Please go through the following assessment criteria for this project.

| **Criteria/Skills** | **Marks** | **Meets Expectations** | **Does Not Meet Expectations** |
| --- | --- | --- | --- |
| Performing time-series analysis | 6 | The code has been replicated successfully and gives an appropriate output in terms of the question and the format | The code has not been replicated successfully and does not give an appropriate output in terms of the question and the format |
| Interpreting time-series analysis | 18 | The required explanations have been given in ample detail and with clarity | The commands and outputs in the Jupyter Notebook have not been explained clearly. |
| Performing clustering analysis | 2 | The code has been replicated successfully and gives an appropriate output in terms of the question and the format | The code has not been replicated successfully and does not give an appropriate output in terms of the question and the format |
| Interpreting clustering analysis | 14 | The required explanations have been given in ample detail and with clarity | The commands and outputs in the Jupyter Notebook have not been explained clearly. |
| Overall | 40 |  |  |

***Please note:***

* *It is suggested that you download the datasets and notebooks and keep working on the project as you go through the course content*
* *The coding portions mentioned in the rubrics above require you to copy and paste the code from elsewhere in the notebook. This has clearly been mentioned in the notebooks as well.*

# Problems & Dataset

## Part 1: Solving for Golden Harvest

Golden Harvest is an upcoming startup that provides customised gold lockets and charms. One of the major problems that it has been facing is associated with the volatility in the price of gold. In order to plan for the financial year better and to look at the possibility of a fixed price offering, they have hired you as a consultant.

You have been provided with the data from the year 1992 to 2019. Here we will be using the data from 1992 to 2017 to build the forecasting model and we will test this model over the data from 2018-2019.

**Dataset & Description**

The dataset to be used for this part of the project is attached below:

**[Dataset: Part 1 –](https://cdn.upgrad.com/uploads/production/01cf4602-bad6-4be5-b8d6-37939168a8c8/gold_price_data.csv" \o "gold_price_data.csv" \t "_blank)** [gold\_Price\_data.csv](https://cdn.upgrad.com/uploads/production/01cf4602-bad6-4be5-b8d6-37939168a8c8/gold_price_data.csv" \o "gold_price_data.csv" \t "_blank)

[file\_download](https://cdn.upgrad.com/uploads/production/01cf4602-bad6-4be5-b8d6-37939168a8c8/gold_price_data.csv" \o "gold_price_data.csv" \t "_blank)**[Download](https://cdn.upgrad.com/uploads/production/01cf4602-bad6-4be5-b8d6-37939168a8c8/gold_price_data.csv" \o "gold_price_data.csv" \t "_blank)**

| **Column** | **Description** |
| --- | --- |
| Date | Month/Day/Year |
| Value | The price of gold at the given date |

## Part 2: Helping Human Resources

Recently, one of your clients (a service-based company) has handled a very complex IT project and has achieved fruitful results. As a token of appreciation, the administration has decided to organise a party for the team that handled the project. However, after interaction with the team members, the HR team found that each person has separate preferences pertaining to food items.

The team came to you to provide a viable solution to the problem at hand. In order to get a better sense of the segments, the employees were given a questionnaire by you and asked to rate the different available options for food items based on their interests.

**Dataset & Description**

The data that was collected is given below:

**[Dataset: Part 2](https://cdn.upgrad.com/uploads/production/935f9aba-2201-4071-aba3-09829ac19864/food_rating.csv" \o "food_rating.csv" \t "_blank)**

**[FoodRating.csv](https://cdn.upgrad.com/uploads/production/935f9aba-2201-4071-aba3-09829ac19864/food_rating.csv" \o "food_rating.csv" \t "_blank)**

[file\_download](https://cdn.upgrad.com/uploads/production/935f9aba-2201-4071-aba3-09829ac19864/food_rating.csv" \o "food_rating.csv" \t "_blank)**[Download](https://cdn.upgrad.com/uploads/production/935f9aba-2201-4071-aba3-09829ac19864/food_rating.csv" \o "food_rating.csv" \t "_blank)**

| **Column** | **Description** |
| --- | --- |
| Emp Id | Unique employee ID |
| Pasta | Rating for pasta as a food option |
| Non-Veg Dish | Rating for non-veg dish as a food option |
| Brownie | Rating for brownie as a food option |
| Lentils | Rating for lentils as a food option |
| Fries | Rating for fries as a food option |
| Fruits | Rating for fruits as a food option |

In the next segment, you will look at the specific questions that need to be answered for this project.

**Assessments & Submission**

**Part 1: Solving for Golden Harvest**

The data scientist in your team has analysed the time-series data and has made predictions on the basis of multiple models. The analysis has been collated in the following notebook:

**[Notebook: Part 1 - AML+Project+Part+1\_TimeSeries+(1).ipynb](https://cdn.upgrad.com/uploads/production/9878026b-90ec-4009-9108-f6a886642807/AML%2BProject%2BPart%2B1_TimeSeries+(1).ipynb" \o "AML+Project+Part+1_TimeSeries (1).ipynb" \t "_blank)**

[file\_download](https://cdn.upgrad.com/uploads/production/9878026b-90ec-4009-9108-f6a886642807/AML%2BProject%2BPart%2B1_TimeSeries+(1).ipynb" \o "AML+Project+Part+1_TimeSeries (1).ipynb" \t "_blank)**[Download](https://cdn.upgrad.com/uploads/production/9878026b-90ec-4009-9108-f6a886642807/AML%2BProject%2BPart%2B1_TimeSeries+(1).ipynb" \o "AML+Project+Part+1_TimeSeries (1).ipynb" \t "_blank)**

In the notebook, you notice that a few of the cell blocks are empty. Given the unavailability of your data scientist at this hour, you decide to fill them in yourself as you go about analysing the outputs in order to make a final decision regarding the predictions.

**Questions (24 Marks):**

***Note:****These questions have been added to the notebook. You need to answer them in the notebook and then submit that for the purpose of evaluation.*

1. Write your observations about the value of gold in detail. **(2 Marks)**
2. According to the results obtained in the previous question, what can you infer about the results obtained? Is the moving average model working fine?  **(2 Marks)**
3. Before we moving ahead building an Autoregressive model, we would be required to remove trend and seasonality from the give time series. Answer, why do trend and seasonality affect Autoregressive models adversely and why is it required to be removed before building an ARIMA model? **(2 Marks)**
4. Explain the significance of the above step(in the notebook). What can we infer from the above results? **(2 Marks)**
5. What is the null hypothesis of the ADF test? What can you infer from the results of this test? **(2 Marks)**
6. Previously we have created functions for plotting rolling mean and standard deviation. Use the same function and check the plot of diff\_data. **(3 Marks)**⇔da. **(3 Marks)**
7. Previously we have created functions for performing adf\_test. Use the same function and check the adf\_test results for diff\_data. **(3 Marks)**
8. Is the data stationary now? Explain the results here. **(2 Marks)**
9. What can you infer from the above plots(in the notebook)? **(2 Marks)**
10. Compare AR, MA and ARMA models using the evaluation metrics and decide the best performing model. **(4 Marks)**

**Part 2: Helping Human Resources**

Your data scientist has now performed clustering analysis on the data gathered through the survey. He has shared the following notebook with you:

**[Notebook: Part 2 - AML+Project+Part+2\_Clustering+(3).ipynb](https://cdn.upgrad.com/uploads/production/59c033c6-0741-4910-b37e-7feb2577d85f/AML+Project+Part+2_Clustering+(3).ipynb" \o "AML Project Part 2_Clustering (3).ipynb" \t "_blank)**

[file\_download](https://cdn.upgrad.com/uploads/production/59c033c6-0741-4910-b37e-7feb2577d85f/AML+Project+Part+2_Clustering+(3).ipynb" \o "AML Project Part 2_Clustering (3).ipynb" \t "_blank)**[Download](https://cdn.upgrad.com/uploads/production/59c033c6-0741-4910-b37e-7feb2577d85f/AML+Project+Part+2_Clustering+(3).ipynb" \o "AML Project Part 2_Clustering (3).ipynb" \t "_blank)**

Again, you notice a couple of empty code-blocks and, while making a note of talking to your decision scientist about these repeated issues, you decide to fill them up as you go through the analysis, drawing relevant insights.

**Questions (16 Marks):**

***Note:****These questions have been added to the notebook. You need to answer them in the notebook and then submit that for the purpose of evaluation.*

1. Do you find any difference in the scale of the variables? Is there an exception? Write your observations from the box-plots. **(2 Marks)**
2. What patterns do you observe here? What insights can you draw from the exploratory data analysis so far? **(3 Marks)**
3. What did we do here(in the previous code block in the notebook)? Why did we do it? **(2 Marks)**
4. According to Hopkins Statistic is there cluster tendency in the data? **(1 Mark)**
5. Write the code for creating the KMeans clusters with the number of clusters = 3. **(2 Marks)**
6. What are the major differences between Employee Segment 0 and Employee Segment 1?**(3 Marks)**
7. Which of the employee segments does not show much interest in any kind of food item? **(3 Marks)**

**Submission**

You need to add your answers to the questions in the respective notebooks at the placeholders mentioned in the notebooks. Once done, you need to zip both the notebooks together and submit.