



Investment Pre-assignment Live Session

Instructor: Sanjay Pandey

About Me

I am working as Data Science Lead at Petpooja with close to 4+ years of experience in Data Science, Deep Learning, predictive Modelling, Machine Learning & Optimization.

I have worked on Recommendation Engine, Face Recogniton, Sentiment Analysis, Computer Vision, Sales Forecasting, OCR, Chatbot.

Built State-of-the-Art in-house speech recognition for Indian accent English language for the company and have deployed the same in a different restaurant for order taking through speech.

Currently working on Automating Invoice Scanning of any domain and understanding its table structure.



Sanjay Pandey

Today's Agenda

- 1 Discussing the Premise of the Assignment The
- 2 objectives for the assignment
- 3 Pre-Modelling Steps
- 4 Analysis Steps and Checkpoints
- 5 Q & A

What is Spark Funds?

You work for Spark Funds, an asset management company. Spark Funds wants to make investments in a few companies. The CEO of Spark Funds wants to understand the global trends in investments so that she can take the investment decisions effectively.

Business objective: The objective is to identify the best sectors, countries, and a suitable investment type for making investments. The overall strategy is to invest where others are investing, implying that the 'best' sectors and countries are the ones 'where most investors are investing'.

Spark Funds has two minor constraints for investments:

1. It wants to invest between **5 to 15 million USD** per round of investment
2. It wants to invest only in **English-speaking countries** because of the ease of communication with the companies it would invest in.

Investment type analysis

Comparing the typical investment amounts in the venture, seed, angel, private equity etc. so that Spark Funds can choose the type that is best suited for their strategy.

Country analysis

Identifying the countries which have been the most heavily invested in the past. These will be Spark Funds' favourites as well.

Sector analysis

Understanding the distribution of investments across the eight main sectors. (Note that we are interested in the **eight 'main sectors'** provided in the mapping file. The file **company** has numerous sub-sector names; hence, you will need to map each sub-sector to its main sector.)

Steps to proceed with the Case Study

There are four major parts that are needed to be done for this case study:

1. Data understanding/Exploration
2. Data cleaning (cleaning missing values, removing redundant columns etc.)
3. Data Analysis
4. Recommendations(Checkpoints)

Data Understanding/Exploration

1. Read the data to Python dataframe
2. Loading data using encoding
 - Try using different encoding formats
 - Use **chardet** library to detect encoding format (Hint:ISO-8859-1)
3. Explore/understand data: `.info()`, `.describe()`, `.head()`, `.tail()`, `.shape`
 - This will give you a sense of what type of dataset you are dealing with
4. Unique Values check: Use of `.unique()` function

- Data Cleaning
 1. Redundant Columns: Use info from Data Exploration steps
 2. Check the percentage of missing values.
 3. Remove all those with very high missing percentage.
 4. For columns with less missing percentage: perform data cleaning steps for both columns and rows
 5. Null Value treatment: Decide if you need to drop null values or impute dummy data into them
 6. Checking out the distribution to impute mean, median or mode.
 7. Check for data consistency to avoid running into issues while merging especially conversion to correct format for strings. Example: convert to lowercase

- Aggregation of data
 - Use `.mean()`, `.median()` type of functions to extract the best average metric
for the purpose
 - Mean vs Median example
 - Join(Merge) Operation to combine dataset from different data frames
 - Avoid use of loops while writing code

Points to remember

- The entire assignment is divided into checkpoints to help you navigate.
- For each checkpoint, you are advised to fill in the tables into the spreadsheet provided in the download segment(Investment.xls).
- Need to submit your insights in a ppt file. Sample PPT is provided. The structure is a suggestion; make sure not to exceed 10 slides.
- Convert the PPT in PDF format for submission. You need to submit a PDF.
- A single ZIP file is needed to be submitted with one Jupyter Notebook, one excel sheet and a PDF file(ppt).
- Don't forget to comment the code properly as it carries separate marks.

Doubts??





Thank
You!