**Market Value Extraction and Companies Extraction**

**Documentation**

1. **Market Value Extraction**

**Aim**: To extract relevant Elements like the CAGR value, projected/forecasted value, forecast year, companies, locations etc, to help the end user, easily find the market related information related to a topic by a quick search. Hence removing the hassle of going over a complete research article.

**Input**: A sentence or a list of sentences, containing market related information.

**Output**: CAGR Value, Forecast Upper and Lower year, Companies, Market segmentation with respect to Countries/Continents.

**Input Format**:

1. **List of sentences**: ['Global paperboard packaging market to reach $236.3 billion by 2027 amid the covid-19 crisis, the global market for paperboard packaging estimated at us$154.4 billion in the year 2020, is projected to reach a revised size of us$236.3 billion by 2027, growing at acagr of 6.3% over the period 2020-2027.', 'Boxboard, one of the segments analyzed in the report, is projected to record 6% cagr and reach us$105.4 billion by the end of the analysis period.', 'Early analysis of the business implications of the pandemic and its induced economic crisis, growth in the containerboard segment is readjusted to a revised 6.4% cagr for the next 7-year period.', 'The u.s. market is estimated at $41.8 billion, while china is forecast to grow the paperboard packaging market in the u.s. is estimated at us$41.8 billion in', 'China, the world`s second largest economy, is forecast to reach a projected market size of us$50.4 billion by the year 2027 trailing a cagr of 9.5% over the analysis period 2020 to 2027.']

OR

1. **Single Sentence**: Global paperboard packaging market to reach $236.3 billion by 2027 amid the covid-19 crisis, the global market for paperboard packaging estimated at us$154.4 billion in the year 2020, is projected to reach a revised size of us$236.3 billion by 2027, growing at acagr of 6.3% over the period 2020-2027.

**Output Format:**

1. **For List of Sentences**:

[

{"cagr":"6.3%","continent":["Global"],"country":["Global"],"current\_value":"$ 154.4 billion","lower\_year":2020,"percentage":"","projected\_value":"$ 236.3 billion","sentence":"global paperboard packaging market to reach $236.3 billion by 2027 amid the covid-19 crisis, the global market for paperboard packaging estimated at us$154.4 billion in the year 2020, is projected to reach a revised size of us$236.3 billion by 2027, growing at acagr of 6.3% over the period 2020-2027.","upper\_year":2027},

{"cagr":"6%","continent":"","country":"","current\_value":"","lower\_year":"","percentage":"","projected\_value":"$ 105.4 billion","sentence":"boxboard, one of the segments analyzed in the report, is projected to record 6% cagr and reach us$105.4 billion by the end of the analysis period.","upper\_year":""},

{"cagr":"9.5%","continent":"","country":["China"],"current\_value":"","lower\_year":2020,"percentage":"","projected\_value":"$ 50.4 billion","sentence":"china, the world`s second largest economy, is forecast to reach a projected market size of us$50.4 billion by the year 2027 trailing a cagr of 9.5% over the analysis period 2020 to 2027.","upper\_year":2027},

{"cagr":"","continent":["Global"],"country":["Global"],"current\_value":"$ 154.4 billion","lower\_year":2020,"percentage":"","projected\_value":"$ 236.3 billion","sentence":"global paperboard packaging market to reach $236.3 billion by 2027 amid the covid-19 crisis, the global market for paperboard packaging estimated at $154.4 billion in the year 2020, is projected to reach a revised size of $236.3 billion by 2027, growing at acagr of 6.3% over the period 2020-2027.","upper\_year":2027},

{"cagr":"","continent":[],"country":["China","U s"],"current\_value":"$ 41.8 billion","lower\_year":"","percentage":"","projected\_value":"$ 41.8 billion","sentence":"the u.s. market is estimated at $41.8 billion, while china is forecast to grow the paperboard packaging market in the u.s. is estimated at $41.8 billion in","upper\_year":""}

]

1. **For Single Sentence**:

[

{"cagr":"6.3%","continent":["Global"],"country":["Global"],"current\_value":"$ 154.4 billion","lower\_year":2020,"percentage":"","projected\_value":"$ 236.3 billion","sentence":" global paperboard packaging market to reach $236.3 billion by 2027 amid the covid-19 crisis, the global market for paperboard packaging estimated at us$154.4 billion in the year 2020, is projected to reach a revised size of us$236.3 billion by 2027, growing at acagr of 6.3% over the period 2020-2027. ","upper\_year":2027}

]

**Process**:

1. Unique count of Percentages and locations are taken. Based on the count, following processes are considered:
   1. If unique percentage count is 1, irrespective of number of unique locations. Rule based approach is considered.

**Rule Based Approach**:

1. Detect whether input contains a list of sentences or just a single sentence.
   1. If list,
      1. Extract relevant sentences having term ‘cagr’ into a list
      2. Extract relevant sentences not having term ‘cagr’ into a list
      3. Extract CAGR/Percentage, Forecast Dates, Market Value, Countries/Continents (As described below) for both group of sentences.
   2. If single sentence,
      1. Detect whether 'cagr' is present in text and group the sentence.
      2. Extract CAGR, Forecast Dates, Market Value, Countries/Continents (As described below)
2. Processes Involved:
   1. Find CAGR/Percentage:
      1. Extract Noun Chunks from SpaCy:
         1. If “%” in a noun chunk -> find CAGR using regex
      2. If CAGR not found:
         1. Detect percentage in SpaCy Entities

Return CAGR/Percentage

* 1. Countries / Continent:
     1. PyCountry:
        1. Check if countries from pycountry library is present in text, amd append them to the list of countries
     2. JSON File:
        1. Check if countries from json file is present in text, and append them to the list of countries/continent
     3. Nationality:
        1. Check if nationalities like Indian, European, Asian, etc are present in text, append relevant country/continent to list
     4. Standardize the countries using country\_converter python library and return list of countries and continent (Eg: “USA” -> “United States”, “US” -> ‘United States’, etc)
  2. Market Value:
     1. Get tokens from SpaCy:
        1. If grammatical filters matched by token (entity type, POS, dependency, etc)
           1. Collect and combine tokens [(“$”, ”193”, “million”) -> “$193 million”
        2. Sort thousands, millions, billions, trillions separately
        3. If number of elements in list:
           1. Number of elements in list > 1 :

Current Value = first element

Projected Value = Last Element

* + - * 1. Number of elements in list == 1 :

If ‘forecast’, ‘reach, 'grow', etc. present in text:

Projected Value = Element

Else

Current Value = Element

Return Projected and Current Value

* 1. Date/Forecast Year:
     1. Get Entities from SpaCy:
        1. If entity type is ‘DATE’ or ‘YEAR’ and 1900 < date < 3000, add date to list
        2. If number of elements in list:
           1. Number of elements in list > 1 :

Sort elements

Lower Year = first element

Upper Year = Last Element

* + - * 1. Number of elements in list == 1 :

If ‘forecast’, ‘reach, 'grow', etc. present in text:

Upper Year = Element

Else

Lower Year = Element

Return Upper Year and Lower Year

1. **Company Extraction**

**Aim**: To extract relevant Ecompanies from market research articles. It helps in finding the key players operating on that research topic.

**Input**: A sentence or a list of sentences, containing companies.

**Output**: Companies.

**Input Format:**

1. **List of Sentences:**

['Looking at the growth prospects of global electric parking brake systems, persistence market research has provided comprehensive insights on the overall likeliness of the business for oem giants in the industry.', 'Likewise, hyundai mobis announced to launch an open innovation center m-cube in silicon valley of the us to make a full-fledged commitment to source and invest in startups possessing new technologies for future such as self-driving vehicles.', 'Europe, east asia and south asia emerging as ultimate destinations for manufacturers the automotive market is estimated to be left with only a handful percentage of new cars on sale that have a manually operated handbrake; dacia and suzuki being the major players in this bracket.']

1. **Single sentence:**

'Tier-ii and other players in the electric parking brake market, namely toshiba corporation, robert bosch gmbh, sanken electric co., ltd., and dura automotive systems, are focusing on introducing cost effective and customized products to improve their market penetration.', 'Persistence market research (pmr) has published a research report on the global electric parking brake market, which serves as summarization of the potential growth of the market.'

**Output Fomat:**

1. **List of Sentences:**

[{'sentence': 'Europe, east asia and south asia emerging as ultimate destinations for manufacturers the automotive market is estimated to be left with only a handful% of new cars on sale that have a manually operated handbrake; dacia and suzuki being the major players in this bracket.', 'companies': ['dacia', 'suzuki']}]

1. **Single Sentence:**

[{'sentence': 'Tier-ii and other players in the electric parking brake market, namely toshiba corporation, robert bosch gmbh, sanken electric co., ltd., and dura automotive systems, are focusing on introducing cost effective and customized products to improve their market penetration.', 'companies': ['robert bosch gmbh', 'sanken electric co., ltd.', 'dura automotive systems']}]

**Process**:

1. Detect if input is list of sentences or single sentence
2. If list:
   1. Select sentences that contain words like ‘key players’, ’companies’, ‘players’, etc.
   2. Pass each sentence to SpaCy to extract entities
      1. If entity type == ‘ORG’
         1. Append to list of companies

If single sentence:

* 1. Pass the sentence to SpaCy to extract entities
     1. If entity type == ‘ORG’
        1. Append to list of companies

**Question Answer Based Approach**:

Question Answer model can be utilized to extract information by asking model questions regarding the input that has been passed.

**Eg**:

**Input text**: ‘Global paperboard packaging market to reach $236.3 billion by 2027 amid the covid-19 crisis, the global market for paperboard packaging estimated at us$154.4 billion in the year 2020, is projected to reach a revised size of us$236.3 billion by 2027, growing at acagr of 6.3% over the period 2020-2027.’

**Process**: