ROKT

Marketing Analyst Skill Test (Level 4)

Background/Problem:

The CRM we use is Hubspot. We have a customer list within Hubspot and each contact has a set of properties associated with it. Tier refers to the size of the opportunity and thus the size of the company. Within the lead form we ask users to provide their job title. We use job titles to inform targeting. This data has never been uniform. We want to understand what titles are most commonly associated with each tier in the decision making process and what business function they exist in. Given your past experience analyzing marketing data please see the attached spreadsheet of dummy CRM data and create a report with insights and explain your approach.

Report

Forethought:

To answer the questions to the problem, we must first gather all our data from various sources. Next, we need to clean and uniform all the datasets before we can analyze them. Finally, with clean, uniformed data as a point of truth, we can visualize the data and provide accurate insights into findings through dashboards and reports.

With that in mind, we also need to be able to scale our process for big datasets, whether they are in thousands, millions, or billions of rows.

Procedures: *changes depending on the types of data and what we want to perform on the data*

1. Initial data clean-up through Excel (Github Repository):

If there are multiple sources of data, SQL queries would be utilized to pull relevant tables together for in-depth analysis.

With the data provided for this prompt (Contact Export), we'll start by cleaning as much data up as possible.

To start out, let's split the data in `Strength - Decision-Making Power / Seniority` so that only A/B/C/D/F/X values are in their own column labeled as `Strength`. The rest of that data's information will be in the `Decision-Making Power/Seniority` column.

We'll do the same for the `Account Tier` column and only keep the numeric value.

Save and export relevant files for future calculations (Copy of Contact - Updated 09172022.csv), making sure to keep the original file.

2. Import updated data into Python for additional data cleaning/uniforming.

We'll continue with further data cleaning by importing the updated .csv file into Jupiter notebook.

Using various python methods and functions, we'll remove empty, non-relevant, non-English, non-descriptive rows, etc. from our dataset.

See below for details of data clean-up/uniforming. We'll focus on `Strength`, `Vertical`, `Job Title`, and `Account Tier` for this clean-up.

We can now start analyzing the data, as our point of truth for accurate visualization, and actionable statistics.

Save and export the file from Jupiter.

3. Finalized data as a point of truth for data visualization: Python/Tableau/DOMO/Power BI/etc.

To see the differences between analyzing raw, not uniformed data compared to finalized data, we can also perform visualization on both datasets. However, we will only explore the filtered data for this assessment.

Now armed with accurate data, we can see which job titles are the most commonly associated with which account tier in the decision-making process and what business function they exist in.

4. Generating a report with detailed dive to answer questions and provide additional insights.

Finally, we can wrap up our findings and any additional insights in reports based on accurate information that can also be automated with time.

A dashboard is a way to gauge the status of a key performance indicator (KPI) while a report is a detailed dive with the goal of answering questions.

For this project, detailed dive is provided throughout all the procedure steps.

Final thoughts:

Overall, this demonstrates the importance of data cleaning. A process that can take weeks, if not months, to achieve data as accurately as possible. Building accurate reports require good data practice and planning so that we're not wasting resources on inaccurate planning. The best part, it can continually scale with data growth.

A total of 20 Hours+ were dedicated to this assessment. I pride myself on being detailed oriented. So while the data cleaning aspect, particularly for `Job Title`, had taken up the most time, I also have learned a lot by testing back and forth various functions as the best path for this project.

Step 2: Import updated data into Python for additional data cleaning/uniforming

Exploring Hubspot Data:

In [1]:
```
from csv import reader
from plotly.subplots import make_subplots
import plotly.graph_objects as go
import pandas as pd
import numpy as np
opened_file = open('data/Copy of Contact - Updated 09172022.csv', encoding="utf8")
read_file = reader(opened_file)
hubspot = list(read_file)
hubspot_header = hubspot[0] # Column names indexing at 0
hubspot = hubspot[1:] # Excluding header
```

In [2]:
```
# Function to easily view the data
def explore_data(dataset, start, end, rows_and_columns=False):
    dataset_slice = dataset[start:end]
    for row in dataset_slice:
        print(row)
        print('\n') # Adds a new (empty) line after each row

    if rows_and_columns:
        print('Number of rows:', len(dataset))
        print('Number of columns:', len(dataset[0]))
```

In [3]:
```
# Test print hubspot dataset
print (hubspot_header)
print ('\n')
explore_data(hubspot, 0, 3, True)
```

```
['Job Title', 'Account Tier', 'Strength', 'Decision-Making Power / Seniority  ', 'Vertic
al', 'Sub-Vertical', 'Pod', 'Industry', 'Original Source', 'Original Source Drill-Down
1']


['EVP, Direct To Consumer', '5', 'A ', 'A - CEO/CXO', 'Retail', 'Fashion', 'Retail', 'Ap
parel & Fashion', 'Offline Sources', 'EXTENSION']


['', '4', 'X ', 'X - Not Specified', 'B2B', 'Software and Digital', 'B2B', 'Internet',
'Offline Sources', 'EXTENSION']


['Senior Technical Consultant', '4', 'X ', 'X - Not Specified', 'B2B', 'Software and Dig
ital', 'B2B', 'Internet', 'Offline Sources', 'EXTENSION']


Number of rows: 887
Number of columns: 10
```

Upon initial assessment of the excel filtered data provided, there are a total of 887 rows with 10 columns. Before we can start analyzing the data, generating accurate reports, and passing the data into business intelligence visualization tools like Power BI/Tableau/etc., we need to clean and uniform the data as much as possible.

Cleaning Hubspot Data

Usually, the process of data cleaning/uniforming takes the longest, as it is a point of truth, and should not be overlooked. For this skill test, I will demonstrate some of the main steps to achieve the most accurate results possible within the given time frame. For this assessment, we are focusing on the following columns: `Strength`, `Vertical`, `Job Title`, and `Account Tier`.

`Strength`

For this column, we have the following values:

```
- A - CEO/CXO
- B - SVP/VP in decision-making position
- C - Lower-level in relevant teams
- D - Lower-level in non-relevant teams
- F - Left the company
- X - Not specified
```

Let's remove all rows that have F and X as they are not helpful or relevant to our data

In [4]:
```
df = pd.DataFrame(hubspot) # Defined our dataframe
strength_df = df[df[2].str.contains("X|F") == False] # Defining new dataframe with Stren

print ('Number of rows:',len(strength_df))
# print (strength_df)
```

```
Number of rows: 704
```

Now our new dataset contains 704 rows and 10 columns. We can calculate the distribution of Strength (A-D) and see where our data is scatttered:

In [5]:
```
# Build a dictionary to store the strength frequency
strength_frequency = {}

for num in strength_df[2]: # Strength index = 2
    if num not in strength_frequency:
        strength_frequency[num] = 1
    else:
        strength_frequency[num] += 1

print(strength_frequency)
```

```
{'A ': 156, 'D ': 206, 'C ': 261, 'B ': 81}
```

From the frequency table, we can see that there are 156 person with the decision making power of CEO/CXO, 81 SVP/VP, 261 lower-level in relevant teams, 206 lower-level in non-relevant teams.

```
- 156 - CEO/CXO
- 081 - SVP/VP in decision-making position
- 261 - Lower-level in relevant teams
- 206 - Lower-level in non-relevant teams
```

`Vertical`

In the `Vertical` column, we have various industries that will provide additional insights. Let's remove rows that have empty values and not useful for this purpose. While we are at it, we can start removing rows with empty values from other relevant columns.

In [6]:
```python
vertical_df = strength_df.copy() # Creating copy of the data
vertical_df[4].replace('', float("NaN"), inplace = True) # Replacing blanks with NaN
vertical_df.dropna(subset = [4], inplace = True) # Remove rows with NaN in Vertical colu

print ('Number of rows:',len(vertical_df))
#print (vertical_df)
```

```
Number of rows: 703
```

After Filtering for `Strength` and `Vertical` columns, our data length is now 703 rows. We will also calculate and see where the distribution of dataset

In [7]:
```python
# Build a dictionary to store the strength frequency
vertical_frequency = {}

for num in vertical_df[4]: # Vertical index = 4
    if num not in vertical_frequency:
        vertical_frequency[num] = 1
    else:
        vertical_frequency[num] += 1

print(vertical_frequency)
```

```
{'Retail': 318, 'Media and Entertainment': 132, 'Food and Beverage': 70, 'Services': 47,
'B2B': 53, 'Ticketing': 37, 'Jobs and Education': 4, 'Travel': 23, 'Sports and Fitness':
13, 'Automotive': 6}
```

We have the following below for the distribution of `Vertical`:

```
- 318 - Retail
- 132 - Media and Entertainment
- 070 - Food and Beverage
- 053 - B2B
- 047 - Services
- 037 - Ticketing
- 023 - Travel
- 013 - Sports and Fitness
- 006 - Automotive
- 004 - Jobs and Education
```

`Job Title`

At first glance, we can see that there are rows within `Job Title` column that are non-English characters, empty filled, non-descriptive, or entered as variations that mean the same thing (e.g. CEO vs. C.E.O vs. Chief Executive Officer).

Instead of using `np.select` to run multiple matches and apply a specific value upon match, we'll utilize a function developed for this purpose by Matt Harrison. The function's description is included below:

In [8]:
```python
def generalize(ser, match_name, default=None, regex=False, case=False):
    ''' Search a series for text matches.
    Based on code from https://www.metasnake.com/blog/pydata-assign.html

    ser: pandas series to search
    match_name: tuple containing text to search for and text to use for normalization
    default: If no match, use this to provide a default value, otherwise use the origina
    regex: Boolean to indicate if match_name contains a  regular expression
    case: Case sensitive search

    Returns a pandas series with the matched value
```

```
    '''
    seen = None
    for match, name in match_name:
        mask = ser.str.contains(match, case=case, regex=regex)
        if seen is None:
            seen = mask
        else:
            seen |= mask
        ser = ser.where(~mask, name)
    if default:
        ser = ser.where(seen, default)
    else:
        ser = ser.where(seen, ser.values)
    return ser
```

This function can be called on a pandas series and expects a list of tuples. The first tuple item is the value to search for and the second is the value to fill in for the matched value. With limited time, we'll focus on some of the titles, not all of them. Those without a match will retain the original value.

In [9]:
```
# Creating a list called title_patterns to store our data
title_patterns = [
                    ('Account Manager', 'Account Manager'), ('Affiliate Manager', 'Affil
                    ('Associate Director', 'Associate Director'), ('Associate Manager',
                    ('Brand Manager', 'Brand Manager'), ('Business Manager','Business Ma
                    ('C.E.O.', 'CEO'), ('Chief Executive Officer', 'CEO'), ('CEO', 'CEO'
                    ('CXO','CEO'), ('Chief Experience Officer','CEO'), ('Chief Operation
                    ('Chief Revenue Officer', 'Chief Revenue Officer'), ('Chief Technolo
                    ('General Manager', 'General Manager'), ('Co-Founder', 'Co-Founder')
                    ('COO', 'COO'), ('Deputy Editor', 'Deputy Editor'), ('Digital Experi
                    ('Digital Marketing Manager', 'Digital Marketing Manager'), ('Direct
                    ('Ecommerce Manager', 'Ecommerce Manager'), ('EVP', 'EVP'), ('Founde
                    ('GM', 'General Manager'), ('Head Of Marketing', 'Head Of Marketing'
                    ('Head of Performance Marketing', 'Head of Performance Marketing'),
                    ('N/a', 'n/a'), ('President', 'President'), ('Developer', 'Developer
                    ('Vice President', 'Vice President'), ('VP', 'Vice President')
                 ]
```

In [10]:
```
# Calling the generalize function
vertical_df['Stored Title'] = generalize(df[0], title_patterns)

print (vertical_df)
```

```
                                            0  1  2  \
0                     EVP, Direct To Consumer  5  A
3              Director of Cinema Intelligence  4  D
4                    Digital Marketing Manager  4  C
8                             Business Manager  6  D
9              Head of Digital Transformation  5  B
..                                        ... .. ..
882                                       CEO  4  A
883               Director, Account Management  5  D
884                     Online Marketing Manager  5  C
885                      Email Marketing Manager  4  C
886  Senior Account Manager - Western Australia  5  D

                                      3                       4  \
0                           A - CEO/CXO                  Retail
3     D - Lower-level in non-relevant teams  Media and Entertainment
4         C - Lower-level in relevant teams  Media and Entertainment
8     D - Lower-level in non-relevant teams                  Retail
9    B - SVP/VP in decision-making position                  Retail
..                                      ...                     ...
882                         A - CEO/CXO               Ticketing
```

```
883   D - Lower-level in non-relevant teams  Media and Entertainment
884       C - Lower-level in relevant teams                   Retail
885       C - Lower-level in relevant teams                   Retail
886   D - Lower-level in non-relevant teams                 Services

                                        5              6                  7  \
0                                 Fashion         Retail  Apparel & Fashion
3          Media and Entertainment - Other  Entertainment      Entertainment
4          Media and Entertainment - Other  Entertainment      Entertainment
8                                 Fashion         Retail           Textiles
9                          Retail - Other         Retail             Retail
..                                    ...            ...                ...
882                                Events  Entertainment      ENTERTAINMENT
883  Media and Entertainment Subscriptions   Subscription             DESIGN
884                               Fashion         Retail             Retail
885                               Fashion         Retail             Retail
886                  Software and Digital          Telco        REAL_ESTATE

                   8                       9                   Stored Title
0    Offline Sources               EXTENSION                 Vice President
3    Offline Sources               EXTENSION                       Director
4    Offline Sources               EXTENSION       Digital Marketing Manager
8    Offline Sources                  IMPORT                Business Manager
9    Offline Sources                  IMPORT  Head of Digital Transformation
..               ...                     ...                            ...
882  Offline Sources                     API                            CEO
883  Offline Sources                     API                       Director
884  Offline Sources                     API        Online Marketing Manager
885  Offline Sources                     API         Email Marketing Manager
886   Organic Search  Unknown keywords (SSL)                 Account Manager

[703 rows x 11 columns]
```

Now that we have an additional column called `Store Title`, we can continue with our analysis:

In [11]:
```python
# Build a dictionary to store the job title frequency
title_frequency = {}

for num in vertical_df['Stored Title']:
    if num not in title_frequency:
        title_frequency[num] = 1
    else:
        title_frequency[num] += 1

#print(title_frequency) # Remove '#' to see the full frequency table
```

`Account Tier`

We already cleaned the `Account Tier` column through Excel, now let's find out how many belong in each level with the latest data:

In [12]:
```python
# Convert dataframe to list
vertical = vertical_df.values.tolist()

# Building a list for each account tier
tier_6 = []
tier_5 = []
tier_4 = []

for row in vertical: # Using the most filtered data vertical_df, account tier index = 1
    tier = row[1]
    if tier == '6':
        tier_6.append(row)
```

```
    elif tier == '5':
        tier_5.append(row)
    else:
        tier_4.append(row)

print ('Count in Tier 6:', len(tier_6))
print ('Count in Tier 5:', len(tier_5))
print ('Count in Tier 4:', len(tier_4))
```

```
Count in Tier 6: 73
Count in Tier 5: 197
Count in Tier 4: 433
```

Now we need to build a frequency table for each tier to see where the job distribution is:

In [13]:
```
#Build a dictionary to store the tier level frequency
tier6_frequency = {}
tier5_frequency = {}
tier4_frequency = {}

for row in tier_6:
    title = row[10]
    if title not in tier6_frequency:
        tier6_frequency[title] = 1
    else:
        tier6_frequency[title] += 1

for row in tier_5:
    title = row[10]
    if title not in tier5_frequency:
        tier5_frequency[title] = 1
    else:
        tier5_frequency[title] += 1

for row in tier_4:
    title = row[10]
    if title not in tier4_frequency:
        tier4_frequency[title] = 1
    else:
        tier4_frequency[title] += 1
```

Back To Top.

Step 3: Finalized data as a point of truth for data visualization: Python/Tableau/DOMO/Power BI/etc.

To understand what titles are most commonly associated with each tier in the decision making process and what business function they exist in, let's visualize our [`Strength`](#cell21), [`Vertical`](#cell22), and [`Account Tier`](#cell23) columns.

Let's dive deeper and take a look at our distribution of various columns using a bar graph and a pie chart with a function that can help us quickly visualize the table and results.

In [14]:
```
# Function to generate frequency tables that show percentages
def freq_table(dataset, index):
    result = {}

    for row in dataset:
        value = row[index]

        if value in result:
```

```python
            result[value] +=1
        else:
            result[value] = 1

    total = sum(result.values())

    for item in result:
        result[item]/=total # Obtain a fraction of the total
        result[item]*=100   # Convert the fraction to a percentage
        result[item] = round(result[item], 2)

    return result


# Another function we can use to display the percentages in a descending order
def display_table(dataset, index=None):

    if isinstance(dataset, list): # If the dataset is a list of lists compute the requir
        dictionary = freq_table(dataset, index)
    else:
        dictionary = dataset # Else treat the dataset as a dictionary

    result = []
    for key, value in dictionary.items():
        result.append((value, key)) # Appends a ('value','key') tuple into results

    result = sorted(result, reverse=True) # Sort the resulting list in descending order
    for item in result:
        print(item[-1], ': ' , item[0])

    return result


# A function that can help us quickly visualize the table and results from the previous
def show_visuals(dataset, index=None, title_a='', title_b = '', main_title='', y_label='

    '''computes analysis tables, then displays Bar and Pie charts obtained from analysis

    # Store the resulting list from calling the display_table function
    item = display_table(dataset, index)

    # Convert the list to a dictionary
    item = dict(item)

    # Assign chart coordinates from the dictionary values
    y_value = list(item.keys())
    x_value = list(item.values())

    # Create a Bar and Pie chart using assigned coordinates
    fig = make_subplots(rows=1, cols=2,
                        specs=[[{"type": "xy"}, {"type": "domain"}]],
                        subplot_titles=(title_a, title_b))

    fig.add_trace(go.Bar(x=x_value[:5],
                          y=y_value[:5],
                          text=y_value,
                          textposition='outside',
                         showlegend=False),row=1,col=1)

    fig.update_yaxes(title_text=y_label, showticklabels=False, row=1, col=1)

    fig.add_trace(go.Pie(labels=x_value,
                          values =y_value,
                         textposition='inside',
                         textinfo='percent+label'), row=1,col=2)
```

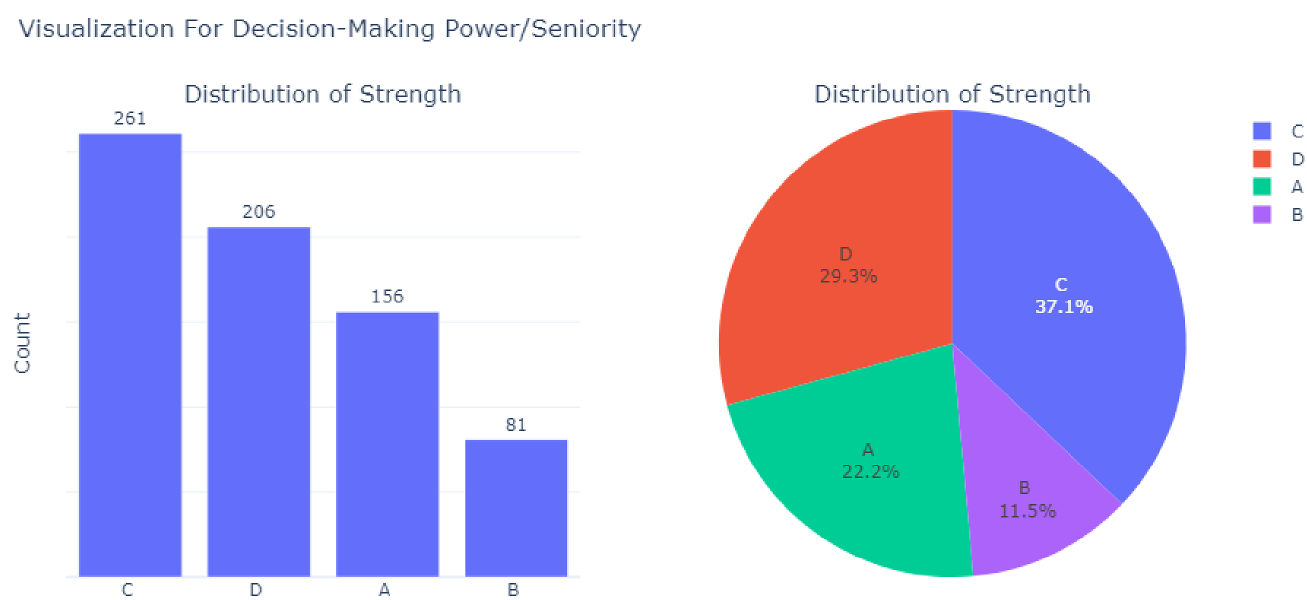
```
    fig.update_layout(template = 'plotly_white', title_text= main_title)

    fig.show('png', width='950')
```

Now that we have everything ready, let's visually explore various column from our dataset, starting with the `Strength` column:

In [15]:
```
# Explore visual data  of Srength index = 2
show_visuals( strength_frequency, 2,
             'Distribution of Strength',
             'Distribution of Strength',
             'Visualization For Decision-Making Power/Seniority ',
             'Count' )
```

```
C  :  261
D  :  206
A  :  156
B  :  81
```



From the charts above, we can see that there are 156 person with the decision making power of CEO/CXO, 81 SVP/VP, 261 lower-level in relevant teams, 206 lower-level in non-relevant teams.

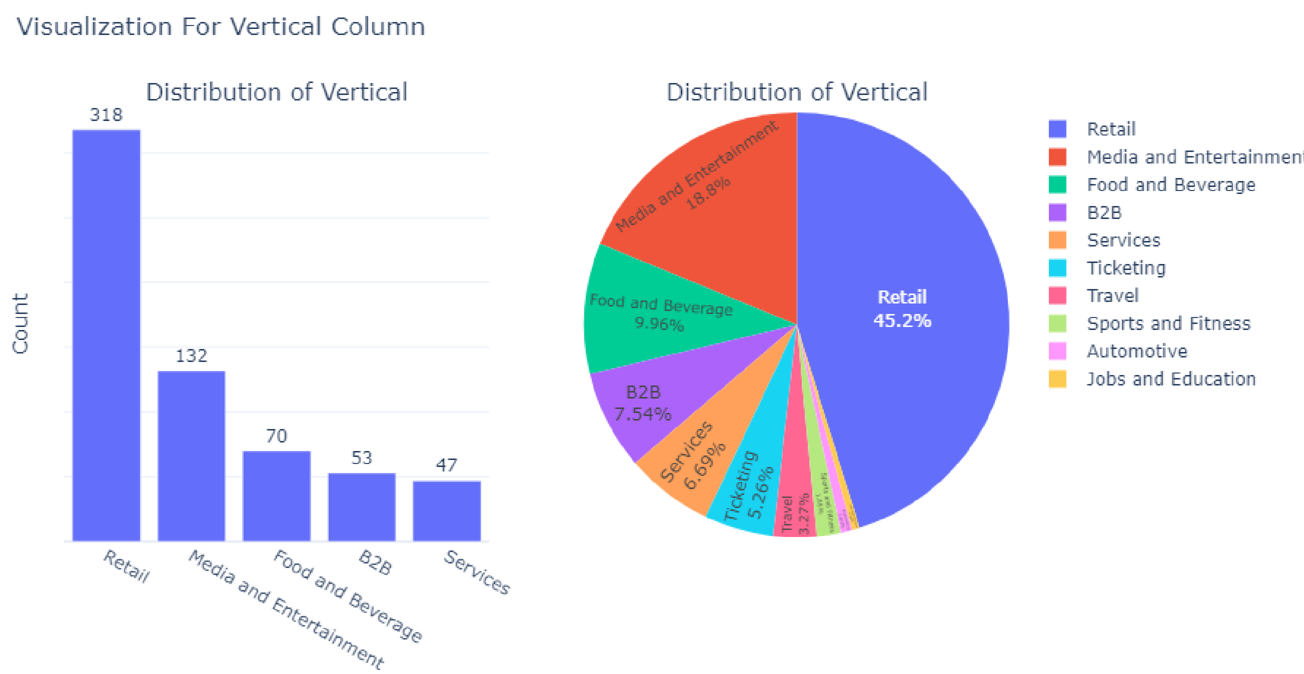
```
- 156 - CEO/CXO
- 081 - SVP/VP in decision-making position
- 261 - Lower-level in relevant teams
- 206 - Lower-level in non-relevant teams
```

Continuing with our insight, let's see the `Vertical` column:

In [16]:
```
# Explore visual data  of Vertical index = 4
show_visuals( vertical_frequency, 4,
             'Distribution of Vertical',
             'Distribution of Vertical',
             'Visualization For Vertical Column',
             'Count' )
```

```
Retail :  318
Media and Entertainment :  132
Food and Beverage :  70
B2B :  53
```

```
Services :  47
Ticketing :  37
Travel :  23
Sports and Fitness :  13
Automotive :  6
Jobs and Education :  4
```



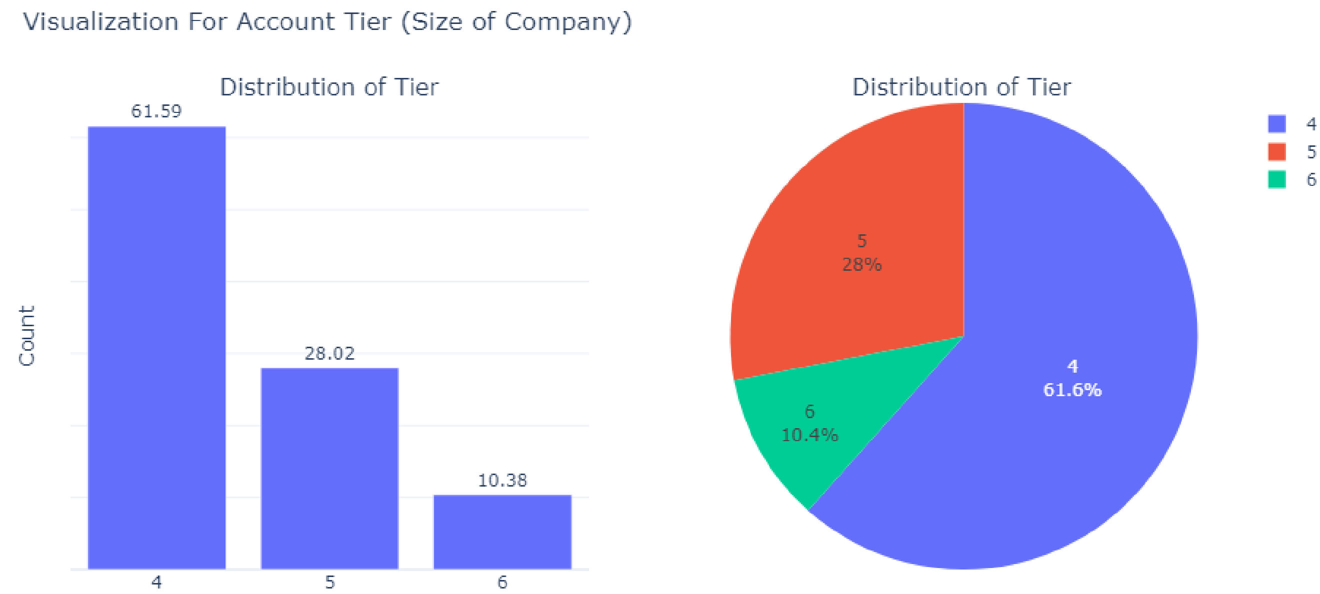
We can see the distribution broken down in the following segment:

```
- 318 - Retail
- 132 - Media and Entertainment
- 070 - Food and Beverage
- 053 - B2B
....and so forth
```

Finally, the `Account Tier` column:

In [17]:
```
# Explore visual data  of Account Tier column, index = 1
show_visuals( vertical, 1,
             'Distribution of Tier',
             'Distribution of Tier',
             'Visualization For Account Tier (Size of Company) ',
             'Count' )
```

```
4 :  61.59
5 :  28.02
6 :  10.38
```



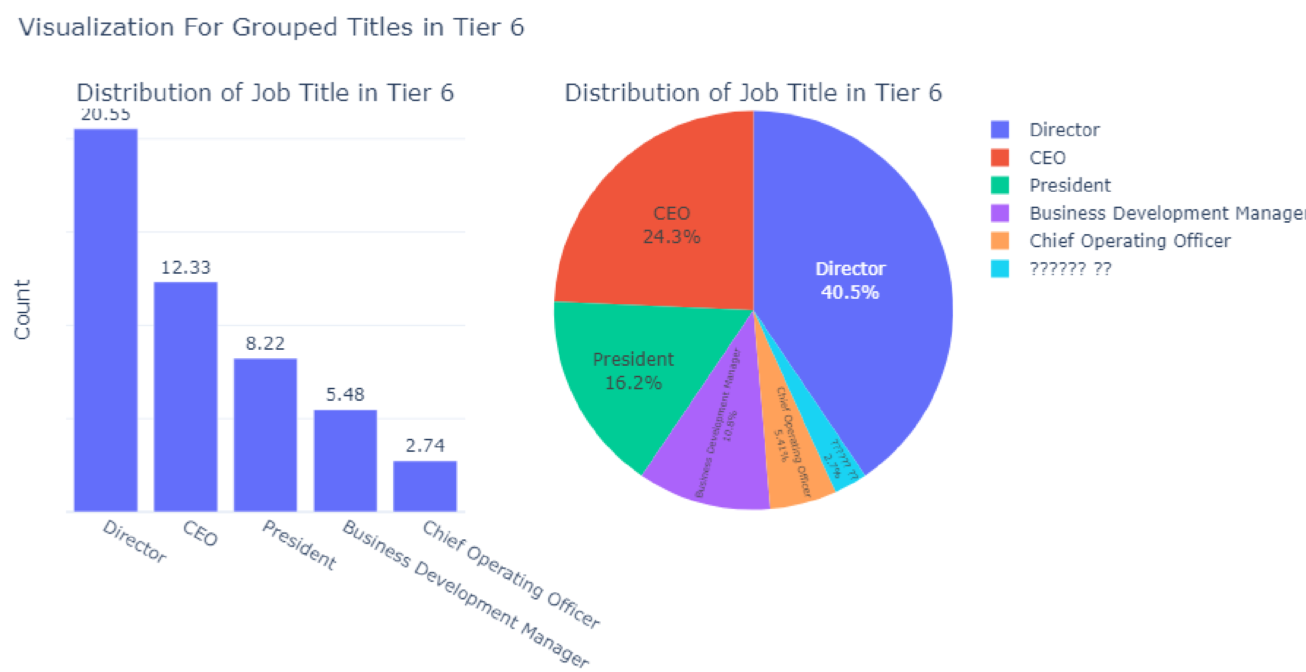
From the charts above, we can see that there are 10.38% of the data is Account Tier 6, 28.02% Account Tier 5, and 61.59% Account Tier 4

```
- 10.38% - Tier 6
- 28.02% - Tier 5
- 61.59% - Tier 4
```

In [18]:
```
# Explore visual data  of Stored Title index = 10
show_visuals( tier_6, 10,
             'Distribution of Job Title in Tier 6',
             'Distribution of Job Title in Tier 6',
             'Visualization For Grouped Titles in Tier 6',
             'Count' )
```

```
Director :  20.55
CEO :  12.33
President :  8.22
Founder :  5.48
Business Development Manager :  5.48
Vice President :  2.74
Chief Operating Officer :  2.74
n/a :  1.37
Tech Lead :  1.37
TBC :  1.37
Support Agent :  1.37
Software Engineer :  1.37
Senior Graphic Designer :  1.37
Projects manager :  1.37
Program Management :  1.37
Patnership Manager :  1.37
Partnerships Manager :  1.37
Marketing & Design Manager :  1.37
Marketing :  1.37
Manager of Strategic Partnerships :  1.37
MD :  1.37
Head of Sales :  1.37
Head of Product :  1.37
Head of Partnerships :  1.37
Head of Ecommerce :  1.37
Head of E-Commerce & Digital Marketing, North America :  1.37
```

```
Head Of Marketing :  1.37
General Manager :  1.37
Executive Sales Manager :  1.37
Ecommerce Systems & Operations Specialist :  1.37
Digital Marketing and Data Entry :  1.37
Digital Marketing Manager :  1.37
Content Designer :  1.37
Chief Financial Officer :  1.37
COO :  1.37
Business Manager :  1.37
Affiliate & Social Manager :  1.37
?????? ?? :  1.37
```

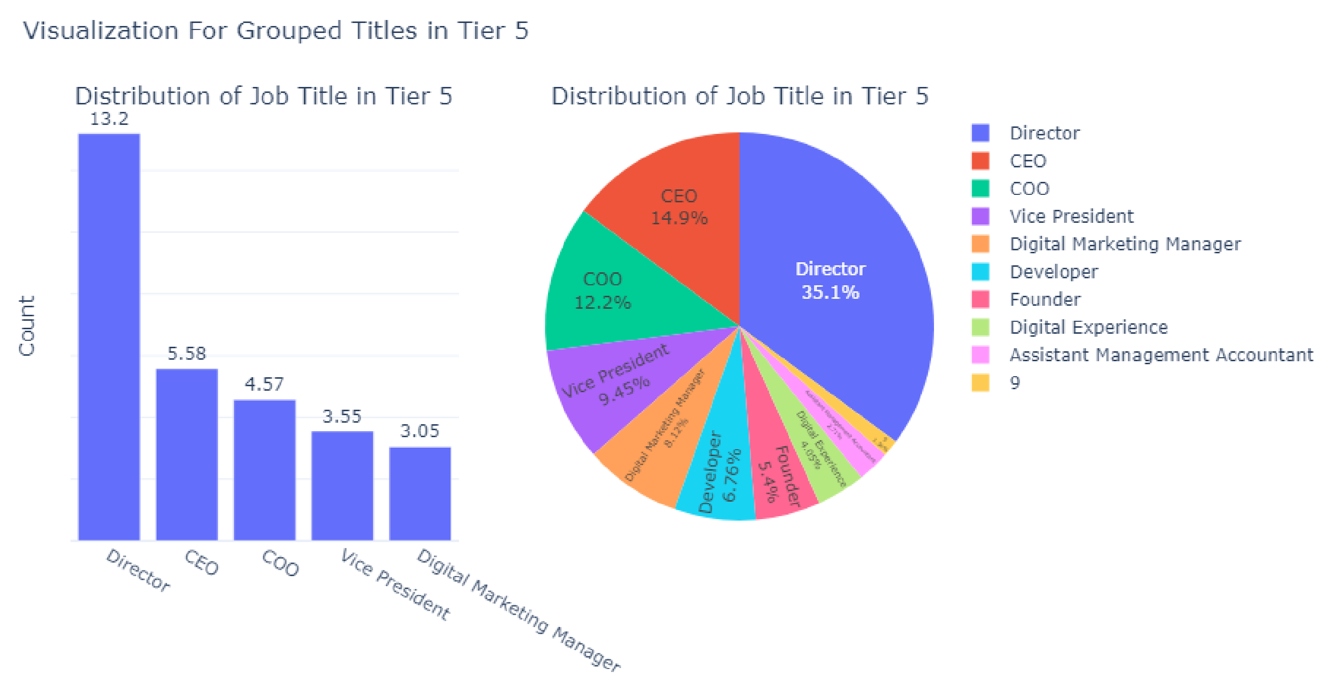


In [19]:
```python
# Explore visual data  of Stored Title index = 10
show_visuals( tier_5, 10,
             'Distribution of Job Title in Tier 5',
             'Distribution of Job Title in Tier 5',
             'Visualization For Grouped Titles in Tier 5',
             'Count' )
```

```
Director :  13.2
President :  5.58
CEO :  5.58
COO :  4.57
Vice President :  3.55
Marketing Manager :  3.05
Digital Marketing Manager :  3.05
Developer :  2.54
Founder :  2.03
Head Of Marketing :  1.52
Digital Experience :  1.52
Product Manager :  1.02
Owner :  1.02
Head of Performance Marketing :  1.02
Head of CRM :  1.02
Financial Controller :  1.02
Finance Manager :  1.02
Chief Technology Officer :  1.02
Business Integration :  1.02
Assistant Management Accountant :  1.02
Web Analytics & Experimentation Strategy Lead :  0.51
Web Analyst :  0.51
User Experience Lead :  0.51
```

Trading Manager : 0.51
Trading Analytics Partner : 0.51
Software Operations Analyst : 0.51
Software Engineer : 0.51
Social Media Specialist : 0.51
Social Media Manager : 0.51
Site Manager : 0.51
Senior Software Engineer : 0.51
Senior Marketing Exec : 0.51
Senior FP&A Manager : 0.51
Senior Customer Advisor : 0.51
Senior CRM Manager, Sony Music- UK : 0.51
Senior CRM Executive - Luxury Division : 0.51
Sales Manager : 0.51
Sales : 0.51
Public Relations Manager : 0.51
Preloved Trading Manager : 0.51
Partnership Marketing Manager : 0.51
Partners and Outreach Manager : 0.51
Partner : 0.51
PPC Manager : 0.51
Operations Specialist : 0.51
Online Marketing Manager : 0.51
National Advertising Sales Manager : 0.51
Media Planner and Buyer : 0.51
Marketing Operations Graduate : 0.51
Marketing Manager, CRM : 0.51
Marketing Communications Manager - eCommerce : 0.51
Marketing - FreeBizMag.com : 0.51
Market Planning and Channel Optimization : 0.51
MD THG Media : 0.51
Information Security Analyst : 0.51
IT Manager : 0.51
Head, eCommerce : 0.51
Head, Social - MyProtein : 0.51
Head, Performance Marketing - Beauty : 0.51
Head of Partnerships : 0.51
Head of Mobile Development : 0.51
Head of Growth : 0.51
Head of Ecommerce and Product : 0.51
Head of Ecommerce : 0.51
Head of Digital Transformation : 0.51
Head of Digital Marketing : 0.51
Head of Business Development : 0.51
Head Of Customer Experience : 0.51
Global Head of Trading - THG Ingenuity : 0.51
General Manager : 0.51
Full Stack Development Lead : 0.51
Finance Officer : 0.51
Email Marketing Executive : 0.51
Ecommerce Trading Manager : 0.51
Ecommerce Manager : 0.51
E-Commerce Dept. : 0.51
Digital Marketing Executive : 0.51
Digital Marketing & Content Manager : 0.51
Digital Leader : 0.51
Digital & Restaurant Experience Marketing Representative : 0.51
Designer : 0.51
Data Intelligence Analyst (Marketing) : 0.51
Data Analyst : 0.51
DIGITAL MARKETING EXECUTIVE : 0.51
Customer Service Manager : 0.51
Customer Relationship Management Lead : 0.51
Creative : 0.51
Chief Operating Officer : 0.51
Chief Marketing Officer : 0.51

```
Chief Growth Officer :  0.51
Chief Financial Officer & Company Secretary :  0.51
Chief Financial Officer :  0.51
Chief Digital Officer :  0.51
Chief Development Officer :  0.51
Chief Data Officer :  0.51
Channel Marketing Manager - Digital Ventures :  0.51
Campaigns and Partnerships Marketing Executive :  0.51
CRM :  0.51
CISO :  0.51
CFO :  0.51
Brand Manager :  0.51
BA :  0.51
Associate Corporate Attorney :  0.51
Affiliate Marketing Manager :  0.51
Affiliate Manager :  0.51
Advertising Executive :  0.51
Account Manager :  0.51
 :  0.51
```



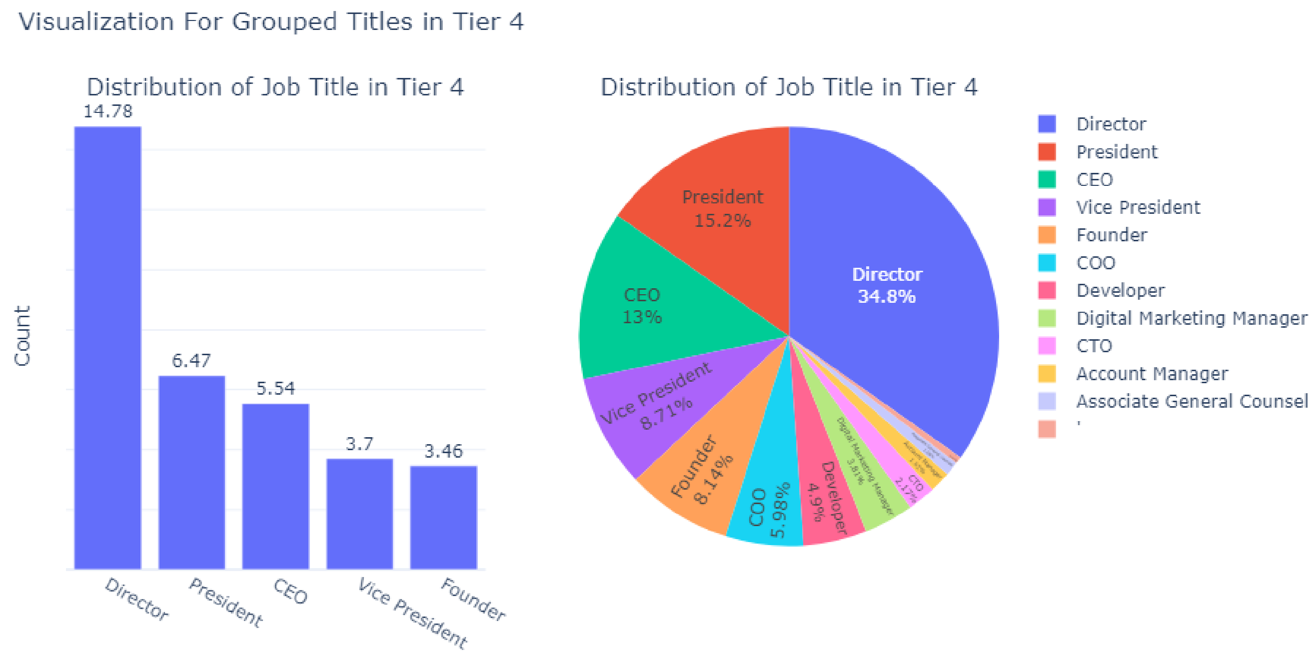
In [20]:
```python
# Explore visual data  of Stored Title index = 10
show_visuals( tier_4, 10,
             'Distribution of Job Title in Tier 4',
             'Distribution of Job Title in Tier 4',
             'Visualization For Grouped Titles in Tier 4',
             'Count' )
```

```
Director :  14.78
President :  6.47
CEO :  5.54
Vice President :  3.7
Founder :  3.46
General Manager :  2.54
COO :  2.54
Marketing Manager :  2.08
Developer :  2.08
Digital Marketing Manager :  1.62
Head Of Marketing :  0.92
Ecommerce Manager :  0.92
Chief Technology Officer :  0.92
Chief Operating Officer :  0.92
Chief Financial Officer :  0.92
CTO :  0.92
```

Software Engineer : 0.69
Senior Product Manager : 0.69
Product Manager : 0.69
National Marketing Manager : 0.69
Email Marketing Manager : 0.69
Deputy Editor : 0.69
Customer Experience Manager : 0.69
CMO : 0.69
Account Manager : 0.69
Tech Lead : 0.46
Senior Software Engineer : 0.46
Senior Legal Counsel : 0.46
Marketing Analyst : 0.46
Growth Marketing Manager : 0.46
Ecommerce Web Project Manager : 0.46
Ecommerce B2C Customer Service Representative : 0.46
Counsel : 0.46
Chief Information Officer : 0.46
Brand Manager : 0.46
Associate Manager : 0.46
Associate General Counsel : 0.46
vm : 0.23
n/a : 0.23
lead in E-commerce : 0.23
eCommerce Specialist : 0.23
eCommerce Sales Representative : 0.23
eCommerce Product Manager : 0.23
eCommerce Business Support Specialist : 0.23
eCommerce Business Analyst : 0.23
e-Commerce Catalog Data Manager : 0.23
Web Engineer : 0.23
Web Development : 0.23
Volunteer : 0.23
V.P., C.M.O. Design and Merchandising : 0.23
UX/UI Designer : 0.23
Technical Team Lead : 0.23
Technical Product Manager : 0.23
Team Leader : 0.23
TEG Financial Accountant : 0.23
TBC : 0.23
Strategic Advisor : 0.23
Staff Software Engineer : 0.23
Sr. Mgr, Retention Marketing : 0.23
Sr. Manager, Growth Marketing : 0.23
Sr. Manager Ecommerce Operations : 0.23
Sr. Manager - Affiliate Marketing : 0.23
Special Projects : 0.23
Social Media Moderator : 0.23
Social Media Content Manager - ANZ : 0.23
Shopify Specialist : 0.23
Seventeen Style Pros : 0.23
Senior Sales Executive : 0.23
Senior Public Relations Executive : 0.23
Senior Marketing Manager : 0.23
Senior Manager of Digital Customization : 0.23
Senior Manager Ecommerce, B2C : 0.23
Senior Engineer, Ecommerce Platform : 0.23
Senior Counsel : 0.23
Senior CRM Manager - Europe : 0.23
Senior Buyer : 0.23
Senior Business Development Manager : 0.23
Senior Analytics Engineer : 0.23
Senior Affiliate Marketing Manager : 0.23
Security Engineer : 0.23
Sales Manager : 0.23
Sales & Marketing Manager : 0.23

Production & Talent Manager : 0.23
Product and Marketing Manager, New Media : 0.23
Product - Online Rewards for Shopping : 0.23
Product : 0.23
Principal Engineering Leader : 0.23
Principal Application Engineer : 0.23
Platform Manager : 0.23
Partnership Manager, Ozsale : 0.23
Partnership Executive : 0.23
Paid Media Senior Associate : 0.23
Paid Media Optimisation Manager : 0.23
Owner : 0.23
Operations Supervisor ECommerce : 0.23
Operations Project Manager : 0.23
Operations Manager : 0.23
Online Merchandising and Content Manager : 0.23
Online Marketing Projects Manager : 0.23
National Partnerships Manager : 0.23
National Marketing Manager (Advertising & Digital) : 0.23
National Festivals Marketing and Palace Cinemas PR Manager : 0.23
National Advertising Manager : 0.23
Mr : 0.23
Marketing Specialist : 0.23
Marketing & Digital Strategy Manager : 0.23
Manager, Search & Display : 0.23
Manager, Global Brand & Content Strategy : 0.23
Manager, Analytics & Insights : 0.23
Manager of Ecommerce Analytics : 0.23
Manager Marketing : 0.23
Manager - Database Marketing and Analytics : 0.23
Manager : 0.23
Logistics Manager : 0.23
Lead Software Engineer : 0.23
Information Security Engineer : 0.23
Head of Sales : 0.23
Head of Product & Strategy : 0.23
Head of Product : 0.23
Head of Investor and Government Relations : 0.23
Head of Finance : 0.23
Head of E-Commerce Marketing : 0.23
Head of Digital, AU + NZ : 0.23
Head of Corporate Development, Strategy, & Exhibitor Relations : 0.23
Head Of Technical Projects : 0.23
Head Of Operations : 0.23
Head Of Finance : 0.23
Head Of Customer Support : 0.23
Head Of Corporate Communications : 0.23
Growth Product Manager : 0.23
Group. Manager Financial Planning and Analysis : 0.23
Group Financial Controller : 0.23
Global IT PMO Lead : 0.23
Global Head of Operations : 0.23
General Counsel, Assistant Secretary, and Interim Chief Operating Officer : 0.23
Full Stack Software Engineer : 0.23
Financial Reporting Manager : 0.23
Financial Controller : 0.23
Financial Analyst : 0.23
Financial Accountant : 0.23
FP&A Analyst : 0.23
Exhibition Marketing Executive : 0.23
Executive Manager of Contracts : 0.23
Executive Digital Editor, Women's Health magazine : 0.23
Enterprise Architect : 0.23
Editorial Administrative Assistant : 0.23
Editor-in-Chief Cosmopolitan : 0.23
Ecommerce Senior Qa Analyst : 0.23

Ecommerce Product Manager : 0.23
Ecommerce Merchandiser : 0.23
E-Commerce Optimisation Manager : 0.23
E-Commerce Operations Manager : 0.23
E-Commerce Manager : 0.23
Display Manager : 0.23
Digital Strategy & Commercial Mgr : 0.23
Digital Product Manager : 0.23
Digital Media Specialist : 0.23
Digital Media Manager : 0.23
Digital Media & Marketing Manager : 0.23
Digital Marketing Specialist : 0.23
Digital Marketing Project Manager : 0.23
Digital Marketing : 0.23
Digital Experience : 0.23
Delivery Lead/Manager : 0.23
Data Manager : 0.23
Customer Support Lead : 0.23
Customer Service Administrator : 0.23
Customer Feedback Specialist : 0.23
Content and Social Media Marketing Manager : 0.23
Content Marketing Manager : 0.23
Commercial Manager : 0.23
Client Success Manager : 0.23
Circulation & Strategy Manager : 0.23
Chief Revenue Officer : 0.23
Chief People Officer : 0.23
Chief Marketing Officer : 0.23
Chief Integration Officer : 0.23
Chief Growth Officer : 0.23
Chief Executive : 0.23
Chief Ecommerce Officer : 0.23
Chief Digital Officer : 0.23
Chief Commercial Officer & MD Global Ticketing : 0.23
CFO : 0.23
Buyer / Category / Supplier Relationship Manager : 0.23
Business Solutions Manager : 0.23
Business Manager : 0.23
Business Development and Finance : 0.23
Business Development Associate : 0.23
Brand Strategist : 0.23
Brand Marketing Manager : 0.23
Brand & Product : 0.23
Australian Marketing Manager : 0.23
Audience Engagement Manager, Women's Health : 0.23
Administrative Assistant, Engineering : 0.23
Acquisition Marketing & Retention : 0.23
Accounts Manager : 0.23
Accountant : 0.23
ANZ Brand & Marketing Manager : 0.23
???????? : 0.23
??????? : 0.23
' : 0.23

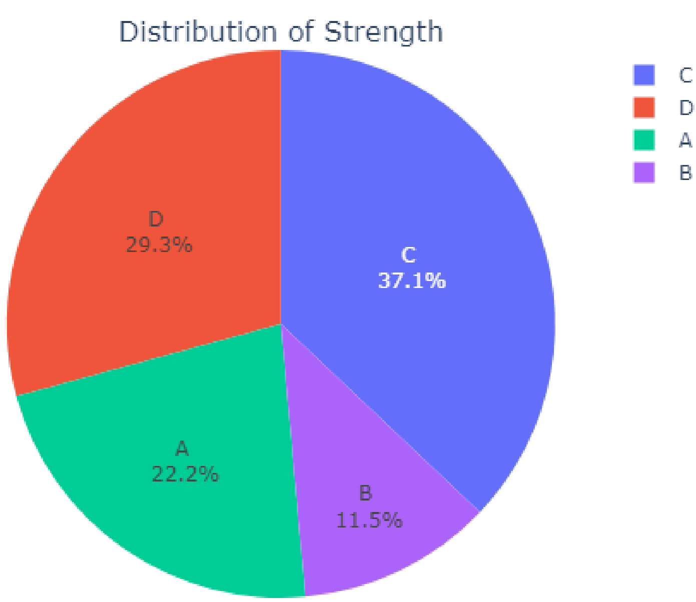


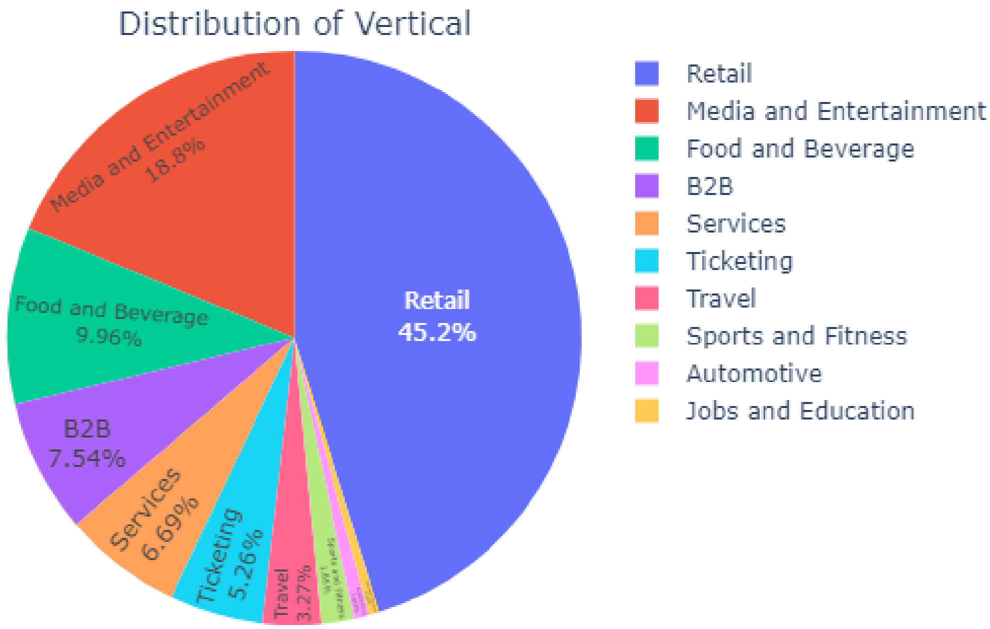
We can see that the distribution of Job titles from the data is aggregated at the Director, President, and CEO levels. With adequate time, we can break the data down further while also much more encompassing, taking into account the variations when users enter the information manually

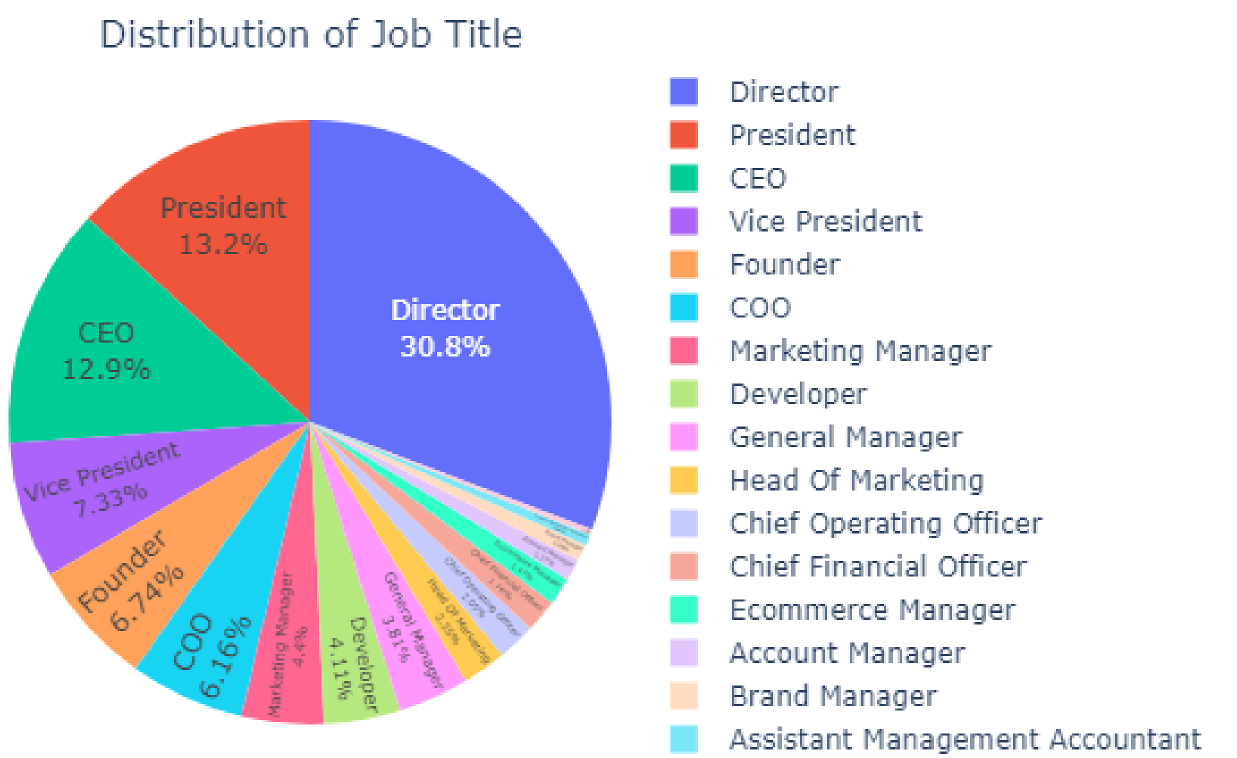
Back To Top.

Conclusion

(Detailed information has been provided throughout this whole report)

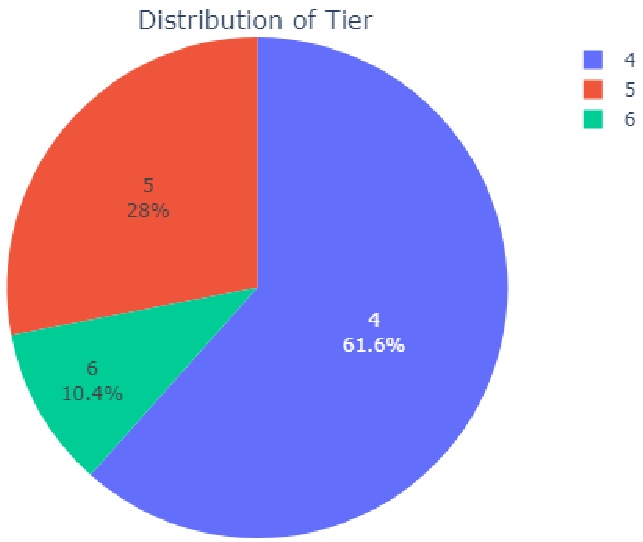


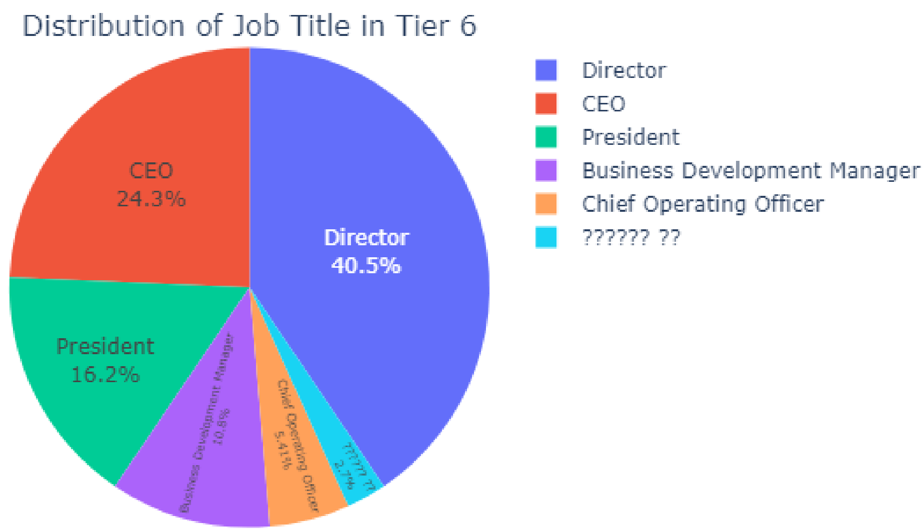


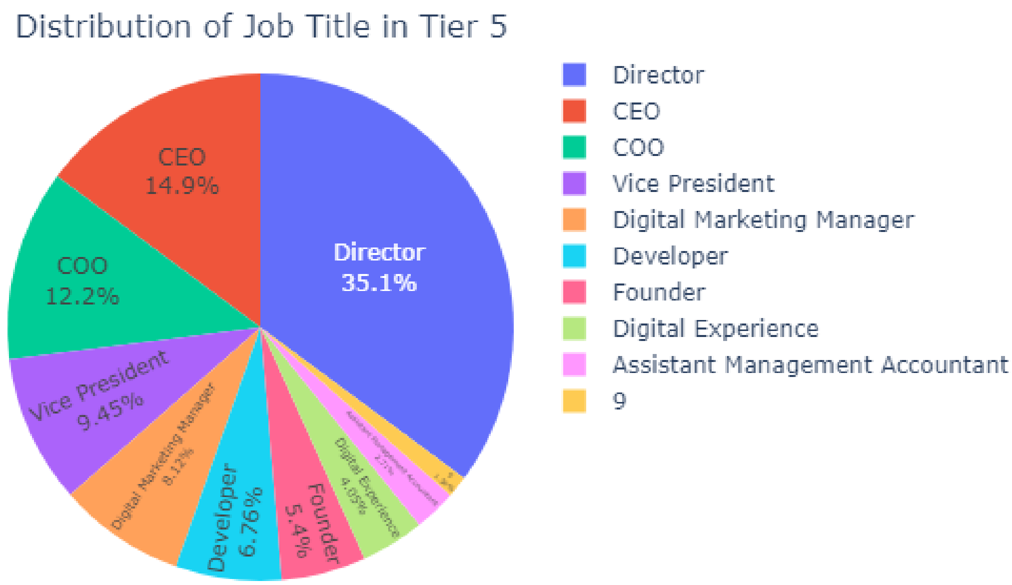


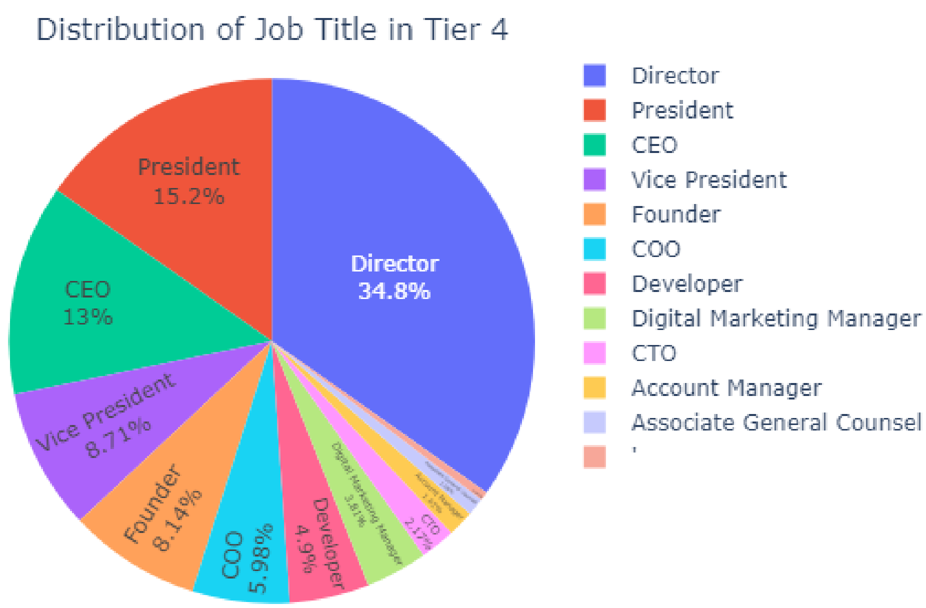
Summary Account Tier, Strength, Vertical, and Grouped Title (not correlated):

Account tier	Strength	Vertical	Grouped Title
Tier 6: 10.38%	CEO/CXO: 156	Retail: 318	Director: 105
Tier 5: 28.02%	SVP/VP: 081	Media and Entertainment: 132	President: 45
Tier 4: 61.59%	Lower-Level (relevant): 261	Food and Beverage: 070	CEO: 44
	Lower-Level (non-relevant): 206	B2B: 053	Vice President: 25
			Founder: 23
			COO: 21


Distribution of Tier
4
5
6
5
28%
4
61.6%
6
10.4%


Distribution of Job Title in Tier 6
Director
CEO
President
Business Development Manager
Chief Operating Officer
?????? ??
CEO
24.3%
Director
40.5%
President
16.2%
Business Development Manager
10.8%
Chief Operating Officer
5.41%
2.7%





Summary of Account Tier (correlated):

Account Tier:	**Tier 6 (10.4%**	**Tier 5 (28%)**	**Tier 4 (61.6%**
	Director: 40.5 %	Director: 35.1%	Director: 34.8%
	CEO: 24.3%	CEO: 14.9%	President: 15.2%
	President: 16.2	COO: 12.2%	CEO: 13%
	Business Development Manager: 10.8%	Vice President: 9.45%	Vice President: 8.71%

We can see *Director* job titles are the most commonly associated with each tier in the the decision making process. *CEO* title is the second most commonly associated for Tier 6 and Tier 5, while it is *President* for Tier 4.

Job Title is an excellent indicator to make a snap judgment for tier in decision-making process. However, using it purely as a metric to determine the audience to target can create lost opportunities. There are many companies with different structures where a role doesn't exist or we can reach out to someone that seems to have the appropriate title, but they are not the decision maker.

Having a company size (Account Tier) with a large number of employees usually mean there are multiple decision makers, while a smaller company will probably only has one or two decision-makers.

Additionally, different industries have different standards and regulations that have varied decision makers.

This showcases how in-depth we can be and how granule the data can get. Further data cleaning should be done to get more accurate results but we will stop right here due to time constrain. Ideally, we need to transform more data, like the `Strength` column from string to numeric so that we can get a better understanding of how it correlates with other values. To quantitively determine the correlation of various metrics, we need to transform things like A/D/C/D to 1/2/3/4 or assign a numeric value that makes sense for our data.

With time constrain, we need to address some of the issues that could provided unreliable information in this project:

```
- Group title can get narrowed down further (in the graph: Cheif Operating
Officer > COO)
- Group any job title that's less than 2 into "Others"
- Non-English titles can get removed with a function to identify ASCII
characters greater than 127
```

Before we're closing this out, let's export our finalized data, a good practice to have it ready for future calculations if needed. Thank you for following my brain crumbs on this journey!

In [21]:
```
# Exporting data to .csv
vertical_df.to_csv('data/Contact_Finalized.csv', index = False)

# Exporting data to excel
vertical_df.to_excel('data/Contact_Finalized.xlsx', index = False)
```

Back To Top.