Forum User Profile

August 27, 2022

1 User Participation in the Odoo Community Forum

This notebook automatically collects data from a user profile using webscraping techniques and based on the collected data it visualizes basic statistics of the participation of this user in the Odoo Community Forum AKA Odoo Help Forum. A pie chart and a calendar heatmap per calendar year (similar to GitHub's contributions plot) are visualization tools employed in this notebook for visualizing the statistics of the user activity.

In order for this notebook to work, the file secret.json must be placed in the same directory as this notebook and it must contain the user credentials required to login into the Odoo Community Forum website (i.e. email and password) and the user ID from the Odoo Community Forum (the user ID can be seen in the URL of the user profile page of the forum).

```
[1]: from datetime import datetime, timedelta
nb_st = datetime.utcnow()
print(f"\nNotebook START time: {nb_st} UTC\n")
```

Notebook START time: 2022-08-27 15:14:55.868153 UTC

```
[2]: %%HTML
     <style>
     Omedia (max-width: 540px) {
       .output .output_subarea {
         max-width: 100%;
       }
     }
     </style>
     <script>
       function code_toggle() {
         if (code_shown){
           $('div.input').hide('500');
           $('#toggleButton').val(' Show Python Code')
         } else {
           $('div.input').show('500');
           $('#toggleButton').val(' Hide Python Code')
         code_shown = !code_shown
```

<IPython.core.display.HTML object>

```
[3]: import pandas as pd
  import numpy as np
  import requests
  from bs4 import BeautifulSoup
  import urllib
  import matplotlib.pyplot as plt
  import seaborn as sns
  import calplot
  import time
  from pathlib import Path
  import json
  import sys
  from contextlib import redirect_stderr
  from IPython.display import clear_output
```

```
[4]: VERBOSE = False
QUICK_DEBUG_RUN = False
```

```
[5]: import importlib

if importlib.util.find_spec('ipywidgets') is not None:
    from tqdm.auto import tqdm

else:
    from tqdm import tqdm

if VERBOSE:
    print("No ipywidgets found, using simple tqdm.")
```

2 Collect data from the user account via webscraping

Based on the User ID and credentials provided in the secret.json login into the user account. Once we are logged into the users account, we have access to all its data. we collect information from the profile and save the collected data locally into the JSON files under the directory named after the User ID in the local data directory.

```
[6]: with open("secret.json", "r") as f:
          s = json.loads(f.read())
      if VERBOSE:
          display({k:s[k] for k in s if k != "password"})
      if (any([k not in s for k in ["user_id", "email", "password"]])
          or any([not s[k] for k in s])):
          raise Exception("Please provide your credentials first!")
     USER_ID = s["user_id"]
 [7]:
     USER_EMAIL, USER_PASSWORD = s["email"], s["password"]
 [8]:
 [9]:
      user_data_dir = Path.cwd() / "data" / str(USER_ID)
[10]: if user_data_dir.is_dir():
          for f in user_data_dir.rglob("*"):
              try:
                  f.unlink()
              except IsADirectoryError:
                  pass
          for f in user_data_dir.rglob("*"):
              f.rmdir()
      else:
          user_data_dir.mkdir(parents=True)
      if VERBOSE:
          print("Local directory of the user data:",
                user_data_dir.relative_to(Path.cwd()))
```

2.1 Login into the User Account

Login into the account and simultaneously go to the profile page by sending a POST request containing the **redirect** key with the desired URI which is undersood by the Odoo backend and it takes us directly to the users profile page.

```
[11]: base_URL = "https://www.odoo.com"
login_URL = urllib.parse.urljoin(base_URL, "/web/login")
```

```
login_payload = {
      'login': USER_EMAIL,
      'password': USER_PASSWORD,
      'redirect': f'/profile/user/{USER_ID}'
      }
[12]: session = requests.Session()
[13]: def raise_on_failure(response, silent=False):
          status_msg = f"HTTP response status code: {response.status_code}"
          if not response.ok:
              raise Exception(status_msg)
          elif VERBOSE and not silent:
              print(status_msg)
          return response
      response = raise_on_failure(session.get(login_URL))
[14]: soup = BeautifulSoup(response.text)
[15]: csrf input = (soup
                    .find("form", {"action": "/web/login", "method": "post"})
                    .find("input", {"name": "csrf token"})
[16]: login_payload['csrf_token'] = csrf_input.get("value")
[17]: if VERBOSE:
          display({k:login_payload[k] for k in login_payload if k != "password"})
[18]: response = raise_on_failure(session.post(login_URL, data=login_payload))
[19]: soup = BeautifulSoup(response.text)
```

2.2 Collect data of User Profile

In this section we are collecting profile data, such as: - Name - Website - City - Country - Biography - Date Joined - Current rank - Current xp - Next rank - xp required for the next rank - Number of Positive votes received - Number of Negative votes received - Avatar image URL - Current rank icon URL - Next rank icon URL - Next rank progress (%)

Data collected in this section is saved locally into the data/"User ID"/user profile.json file.

```
[20]: def add_scheme_to_url(url):
    return urllib.parse.urlunparse(urllib.parse.urlparse(url, scheme="https"))
```

```
[21]: user_profile = {
          "ID": USER_ID,
          "Name": "",
          "Website": "",
          "Email": USER_EMAIL,
          "City": "",
          "Country": "",
          "Biography": "",
      }
[22]: user_profile["Name"] = soup.select(
          ".o_wprofile_header .o_card_people_name")[0].text.strip()
      try:
          user_profile["Website"] = soup.select(
              ".o_wprofile_header i.fa-globe")[0].parent.text.strip()
      except IndexError as e:
          user_profile["Website"] = None
      try:
          location_entries = soup.select(
              ".o wprofile header i.fa-map-marker")[0].find next siblings()
          if len(location entries) > 0:
              user_profile["City"] = location_entries[0].text.strip()
          else:
              user_profile["City"] = None
          if len(location_entries) > 1:
              user_profile["Country"] = location_entries[1].find("span").text.strip()
          else:
              user_profile["Country"] = None
      except IndexError as e:
          user_profile["City"] = None
          user_profile["Country"] = None
      try:
          user_profile["Biography"] = soup.find("h5",
                                                 string="Biography"
                                                ).find_next_sibling()
          if len(user_profile["Biography"].text.strip()) > 0:
              user_profile["Biography"] = str(user_profile["Biography"])
          else:
              user_profile["Biography"] = None
      except Exception as e:
          user_profile["Biography"] = None
```

```
user_profile["Current rank"] = soup.select(
    ".o_wprofile_sidebar .o_wprofile_sidebar_top a")[0].text.strip()
user_profile["Current rank icon"] = add_scheme_to_url(
    soup.select(
        ".o_wprofile_sidebar .o_wprofile_sidebar_top img")[0].get("src"))
user_profile["Joined"] = soup.select(
    "table#o wprofile sidebar table")[0].find("th",
                                              string="Joined"
                                             ).find_next_sibling().text.strip()
(
   user_profile["Positive votes"],
   user_profile["Negative votes"],
) = soup.select(
    "table#o_wprofile_sidebar_table"
)[0].find("th", string="Votes").find_next_sibling().text.strip().split()
(
   user profile["Next rank"],
   user_profile["Current xp"],
   user_profile["Next rank xp"],
) = soup.select(
   "#o wprofile_sidebar_collapse .o_wprofile_progress_circle")[0].text.split()
for key in ("Current xp", "Next rank xp", "Positive votes", "Negative votes"):
   user_profile[key] = int(user_profile[key].replace(",", ""))
user_profile["Next rank progress"] = (
   user_profile["Current xp"] / user_profile["Next rank xp"] * 100
user_profile["Next rank icon"] = add_scheme_to_url(
   soup.select(
        "#o_wprofile_sidebar_collapse .o_wprofile_progress_circle img"
   )[0].get("src")
)
user_profile["Avatar image"] = add_scheme_to_url(
   soup.select_one(
        ".o_wprofile_pict").get("style").split("url(")[1].split(")")[0])
if VERBOSE:
```

```
display(user_profile)
```

2.3 Collect data of Badges

In this section we are collecting data for each badge received by the user: - Badge name - URL of the icon used to for the badge

Data collected in this section is saved locally into the data/"User ID"/user_badges.json file.

```
[24]: user_badges = [{
    "badge_url": add_scheme_to_url(badge.find("img").get("src")),
    "badge_name": badge.text.strip()
}
    for badge in soup.select("#profile_about_badge .card")]

if VERBOSE:
    print(f"\nUser has {len(user_badges)} badges:\n")
    display(user_badges)
```

```
[25]: with open(user_data_dir/"user_badges.json", "w") as f:
    f.write(json.dumps(user_badges))
```

2.4 Collet data of Questions

In this section we are collecting data of questions related to the user, such as: - Questions asked by the user - Questions starred by the user (Favourite Questions) - Questions followed by the user

And the data collected per question is: - URL of the question - Date of question being asked - Votes (total) - Title of the question - Content of the question (i.e. the aqual question)

Data collected in this section is saved locally into the data/"User ID"/related_questions.json file.

```
[26]: related_questions = {}

for question_section in soup.select("#questions >*"):
    section_name = question_section.select_one("h5:first-child").text.strip()
    related_questions[section_name] = []
    question_cards = question_section.select(".card")
```

```
for q in tqdm(question_cards, desc=f'Section of "{section_name}"',
                        total=len(question_cards), dynamic_ncols=True, miniters=1):
              time.sleep(np.random.randint(1, 4)) # rate limit just in case, be kind
              q_url = urllib.parse.urljoin(base_URL, q.select_one("a").get("href"))
              response = raise_on_failure(session.get(q_url), silent=True)
              q_soup = BeautifulSoup(response.text)
              related_questions[section_name].append({
                  "URL": q_url,
                  "time": q soup.select one("article time").text.strip(),
                  "votes": int(q_soup.select_one(".vote_count").text.strip()),
                  "title": q soup.select one("article header").text.strip(),
                  "content": str(
                      q_soup.select_one("article .o_wforum_post_content")),
              })
              if QUICK_DEBUG_RUN:
                  break
          if QUICK_DEBUG_RUN:
              break
      if not VERBOSE:
          clear_output()
[27]: if VERBOSE:
          print("Number of related questions:", len(soup.select("#questions .card")))
[28]: if VERBOSE:
          for section_name in related_questions:
              print(f"\n{section name} ({len(related questions[section name])}):\n")
              if len(related_questions[section_name]) > 0:
                  display(related_questions[section_name][0])
              else:
                  print("This section is empty.")
[29]: with open(user_data_dir/"related_questions.json", "w") as f:
          f.write(json.dumps(related_questions))
```

2.5 Collect data of Answers

In this section we are collecting data of answers posted by the user, per answer we collect: - URL of the answer (which is URL of the question + fragment for the answer) - Date of the answer - Votes (total) of the answer - Whether the answer is accepted or not - Content of the answer (i.e. the aqcual answer) - The following data of the Question addressed by the answer: - Date of the question - Votes (total) of the question - Title of the question - Content of the question (i.e. the aqcual question)

Data collected in this section is saved locally into the data/"User ID"/answers.json file.

```
[30]: answers = []
      answer_cards = soup.select("#answers .card")
      if VERBOSE:
          print("Number of answers:", len(answer_cards))
      for a in tqdm(answer_cards, desc='Answers',
                    total=len(answer_cards), dynamic_ncols=True, miniters=1):
          time.sleep(np.random.randint(1, 4)) # rate limit just in case, be kind
          a_url = urllib.parse.urljoin(base_URL, a.select_one("a").get("href"))
          a html id = "#{}".format(
              urllib.parse.urlparse(a_url).fragment.replace("-","_"))
          response = raise on failure(session.get(a url), silent=True)
          q_soup = BeautifulSoup(response.text)
          a_soup = q_soup.select_one(a_html_id)
          answers.append({
              "URL": a_url,
              "time": a_soup.select_one("time").text.strip(),
              "votes": int(a_soup.select_one(".vote_count").text.strip()),
              "accepted": "o_wforum_answer_correct" in a_soup.attrs["class"],
              "content": str(a_soup.select_one(".o_wforum_readable")),
              "answered question": {
                  "time": q_soup.select_one("article time").text.strip(),
                  "votes": int(q_soup.select_one(".vote_count").text.strip()),
                  "title": q_soup.select_one("article header").text.strip(),
                  "content": str(
                      q_soup.select_one("article .o_wforum_post_content")),
              }
          })
          if QUICK_DEBUG_RUN:
              break
      if not VERBOSE:
          clear_output()
[31]: if VERBOSE:
          print(f"\nAnswers ({len(answers)}):\n")
          if len(answers) > 0:
              display(answers[0])
          else:
              print("This user has not answered any question.")
[32]: with open(user_data_dir/"answers.json", "w") as f:
          f.write(json.dumps(answers))
```

2.6 Collect data of Activity

In this section we are collecting data of the user activity log entries, per entry we collect: - Type of the activity - Time of the activity - URL or the relevant activity

Data collected in this section is saved locally into the data/"User ID"/activity.json file.

```
[35]: with open(user_data_dir/"activity.json", "w") as f:
f.write(json.dumps(activity))
```

2.7 Collect data of Votes Given

display(activity[:10])

if VERBOSE:

In this section we are collecting data of votes given by the user, per vote we collect: - Time of the vote - Whether is the vote positive or negative - The following data is collected for the relevant Question or Answer this vote is givn to: - Title - URL

Data collected in this section is saved locally into the data/"User ID"/votes.json file.

```
[36]: if VERBOSE:
    print(len(soup.select("#votes >div >div")))
```

```
display(votes[:10])

[38]: with open(user_data_dir/"votes.json", "w") as f:
    f.write(json.dumps(votes))
```

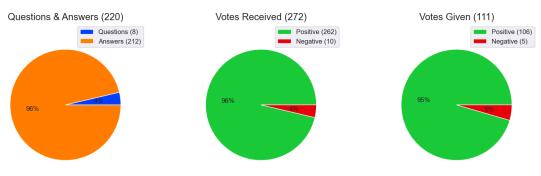
3 Visualize User Participation Data

Pie charts and calendar heatmaps are used here to visualize the following data of the user participation in the forum: - Total number of questions/answers (pie chart) - Total number of positive/negative Votes received (pie chart) - Total number of positive/negative Votes given (pie chart) - Number of questions per day (calendar heatmap) - Number of answers per day (calendar heatmap) - Number of various activities combined per day (calendar heatmap)

3.1 Questions/Answers, Votes Received, Votes Given

Summary of questions/answers, votes received and votes given visualized as pie charts.

```
[39]: plt.show()
      sns.set_theme()
      fig, ax = plt.subplots(ncols=3, figsize=(6*3,9))
      n_questions_asked = len(related_questions["Questions"])
      n_questions_answered = len(answers)
      ax[0].set_title(
          f"Questions & Answers ({n questions asked + n questions answered})",
          fontsize="x-large")
      ax[0].pie([n_questions_asked, n_questions_answered], radius=0.85,
                colors=sns.color_palette('bright')[0:2], autopct='%.0f\%')
      ax[0].legend(labels=[
          f"Questions ({n_questions_asked})",
          f"Answers ({n_questions_answered})",
      ], fontsize="medium", loc="upper right")
      n positive votes received = user profile["Positive votes"]
      n_negative_votes_received = user_profile["Negative votes"]
      ax[1].set title(
          "Votes Received "
          f"({n_positive_votes_received + n_negative_votes_received})",
          fontsize="x-large")
      ax[1].pie([n_positive_votes_received, n_negative_votes_received], radius=0.85,
                colors=sns.color_palette('bright')[2:4],
                autopct='%.0f%%')
      ax[1].legend(labels=[
          f"Positive ({n_positive_votes_received})",
          f"Negative ({n_negative_votes_received})",
      ], fontsize="medium", loc="upper right")
```



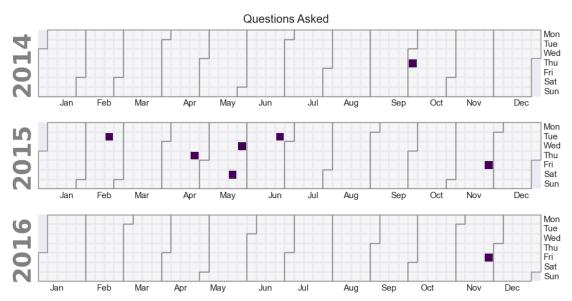
if len(related_question_days[section_name]) > 0:

```
display(related_question_days[section_name][0])
              else:
                  print("This section is empty.")
[43]: answer_days = []
      for answer in answers:
          answer_days.append(datetime.strptime(answer["time"], "%d %B %Y").date())
[44]: if VERBOSE:
          print(f"\nDays of Answers:\n")
          if len(answer_days) > 0:
              display(answer_days[0])
              print("This user has not answered any question.")
[45]: other_activity_days = []
      for a in [x for x in activity if "New" not in x["type"]]:
          other_activity_days.append(
              datetime.strptime(a["time"], "%m/%d/%y, %I:%M %p").date())
[46]: if VERBOSE:
          print(f"\nDays of Activity:\n")
          if len(other_activity_days) > 0:
              display(other_activity_days[0])
          else:
              print("This user has no activity yet.")
[47]: vote_days = []
      for v in votes:
          vote_days.append(
              datetime.strptime(v["time"].split(".")[0], "%Y-%m-%d %H:%M:%S").date())
[48]: if VERBOSE:
          print(f"\nDays of vote:\n")
          if len(vote_days) > 0:
              display(vote_days[0])
              print("This user gave no votes yet.")
```

3.2 Questions Asked

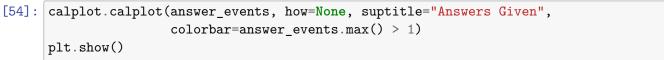
Time series data of number of questions asked per day visualized as a calendar heatmap.

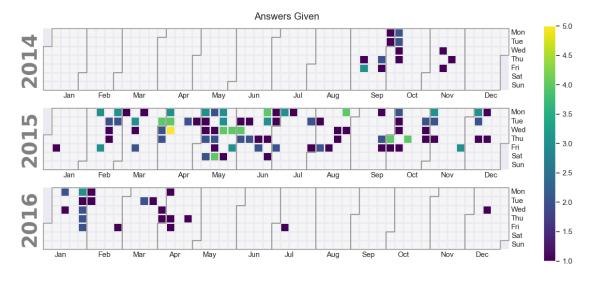
```
[49]: first_question_day = min(related_question_days["Questions"])
      last_question_day = max(related_question_days["Questions"])
      all_question_days = pd.date_range(first_question_day,
                                        last_question_day, freq="D")
      if VERBOSE:
          print(f"Period of questions: {first_question_day} - {last_question_day}")
[50]: | question_events = pd.Series(np.zeros(len(all_question_days)),
                                  index=all_question_days, dtype=int)
      for d in related_question_days["Questions"]:
          question_events.loc[pd.DatetimeIndex([d])] += 1
      if VERBOSE:
          display(question_events)
[51]: with redirect_stderr(io.StringIO()) as f:
          calplot.calplot(question_events, how=None, suptitle="Questions Asked",
                          colorbar=question_events.max() > 1)
      if VERBOSE:
          print(f.getvalue(), file=sys.stderr)
      plt.show()
```



3.3 Answers Given

Time series data of number of answers posted per day visualized as a calendar heatmap.





3.4 Activity

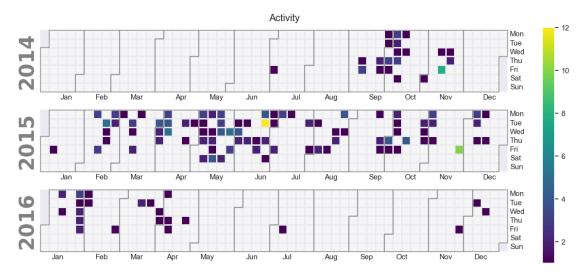
Time series data of number of various types of activity combined per day visualized as a calendar heatmap.

```
[55]: all_day_entries = [
    joined_day,
    *related_question_days["Questions"],
    *answer_days,
```

```
*other_activity_days
]
first_activity_day = min(all_day_entries)
last_activity_day = max(all_day_entries)

all_activity_days = pd.date_range(first_activity_day, last_activity_day, usfreq="D")

if VERBOSE:
    print(f"Period of activity: {first_activity_day} - {last_activity_day}")
```



```
[58]: print(f"\n ** Total Elapsed time: {datetime.utcnow() - nb_st} ** \n")
print(f"Notebook END time: {datetime.utcnow()} UTC\n")
```

** Total Elapsed time: 0:13:03.044926 **

```
[59]: %%capture
      %mkdir OGP_classic
[60]: \%capture
      %%file "OGP_classic/conf.json"
        "base_template": "classic",
        "preprocessors": {
          "500-metadata": {
            "type": "nbconvert.preprocessors.ClearMetadataPreprocessor",
            "enabled": true,
            "clear_notebook_metadata": true,
            "clear_cell_metadata": true
          },
          "900-files": {
            "type": "nbconvert.preprocessors.ExtractOutputPreprocessor",
            "enabled": true
          }
        }
      }
[61]: | %%capture
      %%file "OGP_classic/index.html.j2"
      {%- extends 'classic/index.html.j2' -%}
      {%- block html_head -%}
      {# OGP attributes for shareability #}
      <meta property="og:url"</pre>
                                        content="https://sentinel-1.github.io/
       →odoo_forum_user_profile/" />
      <meta property="og:type"</pre>
                                       content="article" />
      <meta property="og:title"</pre>
                                        content="User Participation in the Odoo⊔

→Community Forum" />
      <meta property="og:description" content="Activity statistics visualized in a_</pre>
       ⇔way similar to GitHub's contributions plot" />
      <meta property="og:image"</pre>
                                        content="https://raw.githubusercontent.com/
       sentinel-1/odoo_forum_user_profile/master/images/OdooProfileScreenshot.png" /
      <meta property="og:image:alt"</pre>
                                        content="Odoo Community Forum Profile_
       ⇔Screenshot" />
      <meta property="og:image:type" content="image/png" />
      <meta property="og:image:width" content="1302" />
      <meta property="og:image:height" content="987" />
      <meta property="article:published_time" content="2022-08-20T09:59:43+00:00" />
```

```
<meta property="article:modified_time"</pre>
                                         content="{{ resources.
 siso8610_datetime_utcnow }}" />
<meta property="article:publisher"</pre>
                                         content="https://sentinel-1.github.io" /
                                         content="https://github.com/sentinel-1"
<meta property="article:author"</pre>
⇔/>
<meta property="article:section"</pre>
                                         content="datascience" />
<meta property="article:tag"</pre>
                                         content="datascience" />
<meta property="article:tag"</pre>
                                         content="Python" />
<meta property="article:tag"</pre>
                                         content="data" />
<meta property="article:tag"</pre>
                                         content="timeseries" />
<meta property="article:tag"</pre>
                                         content="analytics" />
<meta property="article:tag"</pre>
                                         content="datavisualization" />
<meta property="article:tag"</pre>
                                         content="bigdataunit" />
<meta property="article:tag"</pre>
                                         content="visualization" />
<meta property="article:tag"</pre>
                                         content="webscraping" />
<meta property="article:tag"</pre>
                                         content="odoo" />
<meta property="article:tag"</pre>
                                         content="forum" />
<meta property="article:tag"</pre>
                                         content="user" />
<link rel="icon" type="image/x-icon" href="../favicon.ico">
{{ super() }}
{%- endblock html_head -%}
{% block body header %}
<body>
<div class="container">
  <nav class="navbar navbar-default">
    <div class="container-fluid">
      <a href="/">
            <svg xmlns="http://www.w3.org/2000/svg"</pre>
                 viewBox="0 0 576 512" width="1em">
              <path
                fill="#999999"
d="M 288,0 574,288 511,288 511,511 352,511 352,352 223,352 223,511 62,511
 ⇔64,288 0,288 Z"
              />
            </svg> Home
          </a>
```

```
role="presentation" class="active">
         <a href="/odoo_forum_user_profile/"> English </a>
       <a href="/odoo_forum_user_profile/ka/">
                                                  </a>
       </div>
 </nav>
</div>
 <div tabindex="-1" id="notebook" class="border-box-sizing">
   <div class="container" id="notebook-container">
{% endblock body_header %}
{% block body_footer %}
   </div>
 </div>
 <footer>
   <div class="container"</pre>
        style="display:flex; flex-direction: row; justify-content: center; __
 →align-items: center;">
      © 2022
       <a href="https://github.com/sentinel-1" target="_blank">Sentinel-1</a>
     <!-- TOP.GE ASYNC COUNTER CODE -->
     <div id="top-ge-counter-container" data-site-id="116052"</pre>
          style="margin-right: 3.7em; float: right; "></div>
     <script async src="//counter.top.ge/counter.js"></script>
     <!-- / END OF TOP.GE COUNTER CODE -->
     <!-- ANALYTICS.LAGOGAL.COM -->
     <div id="analytics-lagogal-com-access" data-site-id="20221"</pre>
          style="margin: 0;padding: 0;"></div>
     <script async src="//analytics.lagogal.com/access.js"></script>
     <!-- / END OF ANALYTICS.LAGOGAL.COM -->
    </div>
 </footer>
</body>
{% endblock body_footer %}
```

This notebook is originally published under the Apache License (Version 2.0) at the following GitHub repository: $sentinel-1/odoo_forum_user_profile$

For the issues, feedback or suggestions regarding the original notebook (if any) feel free to open an

issue at the corresponding Issues page of the repository