**import** **os**

os.environ["OPENAI\_API\_KEY"] = "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"

**from** **langchain.agents** **import** create\_csv\_agent

**from** **langchain.llms** **import** OpenAI

agent\_new = create\_csv\_agent(OpenAI(temperature=0), 'pracwork\_new.csv', verbose=**True**)

agent\_old = create\_csv\_agent(OpenAI(temperature=0), 'pracwork\_old.csv', verbose=**True**)

In [23]:

agent\_new.run("Make University Of Auckland and The University Of Auckland the same company")

**> Entering new AgentExecutor chain...**

**Thought: I need to find the rows with the company name University Of Auckland**

**Action: python\_repl\_ast**

**Action Input: df[df['Company Name'] == 'University Of Auckland']**

Observation:  **ID Student Name Discipline Company Name \**

**33 249898671 Michael Vaeono CIV University Of Auckland**

**73 313816905 William Hong ELEC University Of Auckland**

**111 676161666 Sophia Wei CMMT University Of Auckland**

**138 745626843 Amanda Lee ENGS University Of Auckland**

**151 505321426 Paul Cha ENGS University Of Auckland**

**... ... ... ... ...**

**4016 863813463 Helen Huang CSYS University Of Auckland**

**4020 710513670 Jennifer Lu BIME University Of Auckland**

**4040 364969506 Tyler Kim SOFT University Of Auckland**

**4041 329128298 Grun Wua Wong CSYS University Of Auckland**

**4048 995049364 Rory Reade MECT University Of Auckland**

**Company Address Company Country \**

**33 20 Symonds Street, Auckland Cbd, Auckland 1010 New Zealand (NZ)**

**73 314-390 Khyber Pass Road, Newmarket, Auckland ... New Zealand (NZ)**

**111 20 Symonds Street New Zealand (NZ)**

**138 23 Symonds Street New Zealand (NZ)**

**151 70 Symonds Street, Grafton New Zealand (NZ)**

**... ... ...**

**4016 The University Of Auckland\nPrivate Bag 92019\... New Zealand (NZ)**

**4020 70 Symonds Street, City Campus New Zealand (NZ)**

**4040 12 Grafton Road New Zealand (NZ)**

**4041 20 Symonds Street, Auckland Cbd New Zealand (NZ)**

**4048 Auckland Cbd, Auckland 1010 New Zealand (NZ)**

**Company Website Supervisor \**

**33 https://www.auckland.ac.nz/en/engineering.html Steve Roberts**

**73 https://www.auckland.ac.nz/en.html Jackman Lin**

**111 https://www.auckland.ac.nz/en/engineering.html Josh Workman**

**138 https://www.auckland.ac.nz Heloise Stevance**

**151 https://www.auckland.ac.nz/en.html Andrew Mason**

**... ... ...**

**4016 https://www.auckland.ac.nz/en.html Ho Seok Ahn**

**4020 https://www.auckland.ac.nz/en.html?gclid=cj0kc... Ho Seok Ahn**

**4040 not available Sri Nagappan**

**4041 https://www.auckland.ac.nz/en.html Ho Seok Ahn**

**4048 https://www.auckland.ac.nz/en.html Rebecca Wallace**

**Supervisor Phone Supervisor Email Work From \**

**33 210701474 s.roberts@auckland.ac.nz 11/25/2020**

**73 022 622 6061 jackman.lin@auckland.ac.nz 12/01/2020**

**111 2102796004 j.workman@auckland.ac.nz 12/01/2020**

**138 99238136 heloise.stevance@auckland.ac.nz 12/07/2020**

**151 09 923 7909 a.mason@auckland.ac.nz 11/25/2020**

**... ... ... ...**

**4016 006493737599 Ext.87860 hs.ahn@auckland.ac.nz 7/11/2022**

**4020 6493737599 hs.ahn@auckland.ac.nz 7/07/2022**

**4040 0 s.nagappan@auckland.ac.nz 2/25/2021**

**4041 0064 9 923 7860 hs.ahn@auckland.ac.nz 12/04/2019**

**4048 276062971 rebecca.wallace@auckland.ac.nz 2/14/2022**

**Work To Work Type Registration Date**

**33 2/25/2021 Not Available 3/15/2022**

**73 3/01/2021 Sub-Professional 12/02/2020**

**111 2/24/2021 Sub-Professional 2/22/2021**

**138 2/24/2021 Sub-Professional 2/14/2021**

**151 2/24/2021 Sub-Professional 2/24/2021**

**... ... ... ...**

**4016 7/15/2022 General 7/25/2022**

**4020 7/21/2022 General 7/25/2022**

**4040 10/22/2021 General 3/17/2022**

**4041 2/28/2020 General 8/28/2020**

**4048 2/10/2023 Not Available 4/27/2022**

**[131 rows x 14 columns]**

Thought: **I need to update the Company Name column to The University Of Auckland**

**Action: python\_repl\_ast**

**Action Input: df.loc[df['Company Name'] == 'University Of Auckland', 'Company Name'] = 'The University Of Auckland'**

Observation:

Thought: **I now know the final answer**

**Final Answer: The Company Name column has been updated to The University Of Auckland for all rows with the Company Name University Of Auckland.**

**> Finished chain.**

Out[23]:

'The Company Name column has been updated to The University Of Auckland for all rows with the Company Name University Of Auckland.'

In [4]:

agent\_new.run("Count number of projects by company give me the top 10")

**> Entering new AgentExecutor chain...**

**Thought: I need to group the data by company and count the number of projects**

**Action: python\_repl\_ast**

**Action Input: df.groupby('Company Name').count()['ID'].sort\_values(ascending=False).head(10)**

Observation: **Company Name**

**The University Of Auckland 199**

**Fisher And Paykel Healthcare 74**

**Fisher Paykel Healthcare 72**

**Beca 48**

**Auckland Transport 43**

**Auckland Bioengineering Institute 36**

**Downer 29**

**Downer Nz 25**

**Aurecon 24**

**Tonkin Taylor 22**

**Name: ID, dtype: int64**

Thought: **I now know the final answer**

**Final Answer: The top 10 companies with the most projects are The University Of Auckland, Fisher And Paykel Healthcare, Fisher Paykel Healthcare, Beca, Auckland Transport, Auckland Bioengineering Institute, Downer, Downer Nz, Aurecon, and Tonkin Taylor.**

**> Finished chain.**

Out[4]:

'The top 10 companies with the most projects are The University Of Auckland, Fisher And Paykel Healthcare, Fisher Paykel Healthcare, Beca, Auckland Transport, Auckland Bioengineering Institute, Downer, Downer Nz, Aurecon, and Tonkin Taylor.'

In [24]:

agent\_new.run("Turn any company with Fisher + Paykel + Healthcare in their name into Fisher and Paykel Healthcare the same company")

**> Entering new AgentExecutor chain...**

**Thought: I need to find any companies with Fisher + Paykel + Healthcare in their name**

**Action: python\_repl\_ast**

**Action Input: df[df['Company Name'].str.contains('Fisher + Paykel + Healthcare')]**

Observation: **ValueError: Cannot mask with non-boolean array containing NA / NaN values**

Thought: **I need to use a different approach**

**Action: python\_repl\_ast**

**Action Input: df[df['Company Name'].str.contains('Fisher') & df['Company Name'].str.contains('Paykel') & df['Company Name'].str.contains('Healthcare')]**

Observation:  **ID Student Name Discipline Company Name \**

**43 142167246 Victoria Skeggs SOFT Fisher And Paykel Healthcare**

**59 541432211 Wenxuan Lu ELEC Fisher And Paykel Healthcare**

**67 6693979 Carl Velasco MECT Fisher Paykel Healthcare**

**69 694475730 Aaryaman Desai MECT Fisher And Paykel Healthcare**

**112 981867593 Rishav Narayan MECH Fisher Paykel Healthcare**

**... ... ... ... ...**

**3875 286933069 Nikhil Kumar CSYS Fisher And Paykel Healthcare**

**3939 584332560 Ben Sharp BIME Fisher And Paykel Healthcare**

**3943 649574844 Benjamin Salmon SOFT Fisher Paykel Healthcare**

**3974 750738161 Katherine Luo CSYS Fisher And Paykel Healthcare**

**4046 876469374 Ella Dixon BIME Fisher And Paykel Healthcare**

**Company Address Company Country \**

**43 15 Maurice Paykel Place, East Tamaki New Zealand (NZ)**

**59 15 Maurice Paykel Place, East Tmaki, Auckland ... New Zealand (NZ)**

**67 15 Maurice Paykel Place, East Tmaki, Auckland ... New Zealand (NZ)**

**69 15 Maurice Paykel Place, East Tmaki, Auckland ... New Zealand (NZ)**

**112 15 Maurice Paykel Place, East Tmaki, Auckland ... New Zealand (NZ)**

**... ... ...**

**3875 15 Maurice Paykel Place New Zealand (NZ)**

**3939 15 Maurice Paykel Place, East Tamaki, Auckland... New Zealand (NZ)**

**3943 15 Maurice Paykel Place East Tamaki New Zealand (NZ)**

**3974 15 Maurice Paykel Place, East Tamaki New Zealand (NZ)**

**4046 15 Maurice Paykel Place New Zealand (NZ)**

**Company Website Supervisor \**

**43 https://www.fphcare.com Sarah Lowe**

**59 https://www.fphcare.com/nz/ Rolando Bagunu**

**67 https://www.fphcare.com/nz/ Fernando Zapata**

**69 https://www.fphcare.com/nz/ Ruben Tan**

**112 https://www.fphcare.com/nz/ Peter Graham**

**... ... ...**

**3875 https://www.fphcare.com/nz/ Vaspan Raimalwala**

**3939 https://www.fphcare.com/nz/ James Gordon**

**3943 https://www.fphcare.com/nz/ George Stirling**

**3974 https://www.fphcare.com/nz/ Hemita Sami**

**4046 https://www.fphcare.com/nz/ Mike Ronayne**

**Supervisor Phone Supervisor Email Work From \**

**43 095740123x81383 sarah.lowe@fphcare.co.nz 11/19/2018**

**59 095740100 ext. 6132 rolando.bagunu@fphcare.co.nz 11/29/2021**

**67 21512042 fernando.zapata@fphcare.co.nz 11/21/2016**

**69 223067152 ruben.tan@gmail.com 11/20/2021**

**112 0064 9 574 0100 EXT 8535 peter.graham@fphcare.co.nz 11/22/2021**

**... ... ... ...**

**3875 95740123 vaspan.raimalwala@fphcare.co.nz 8/22/2022**

**3939 NaN james.gordon@fphcare.co.nz 11/22/2021**

**3943 64274266547 george.stirling@fphcare.co.nz 11/22/2021**

**3974 210304146 h\_chand@hotmail.com 9/13/2021**

**4046 095740123 ext 8745 michael.ronayne@fphcare.co.nz 11/18/2019**

**Work To Work Type Registration Date**

**43 3/01/2019 General; Sub-Professional 8/17/2020**

**59 2/25/2022 General 2/23/2022**

**67 2/24/2017 General 3/21/2021**

**69 2/25/2022 Sub-Professional 8/08/2022**

**112 2/25/2022 Sub-Professional 2/18/2022**

**... ... ... ...**

**3875 12/15/2022 Not Available 8/24/2022**

**3939 2/25/2022 Sub-Professional 12/02/2021**

**3943 2/25/2022 Not Available 2/25/2022**

**3974 4/01/2022 General 9/04/2022**

**4046 2/28/2020 Sub-Professional 2/24/2020**

**[154 rows x 14 columns]**

Thought: **I need to update the company name**

**Action: python\_repl\_ast**

**Action Input: df.loc[df['Company Name'].str.contains('Fisher') & df['Company Name'].str.contains('Paykel') & df['Company Name'].str.contains('Healthcare'), 'Company Name'] = 'Fisher and Paykel Healthcare'**

Observation:

Thought: **I now know the final answer**

**Final Answer: All companies with Fisher + Paykel + Healthcare in their name have been updated to Fisher and Paykel Healthcare.**

**> Finished chain.**

Out[24]:

'All companies with Fisher + Paykel + Healthcare in their name have been updated to Fisher and Paykel Healthcare.'

In [25]:

agent\_new.run("Count number of projects by company give me the top 10")

**> Entering new AgentExecutor chain...**

**Thought: I need to group the data by company and count the number of projects**

**Action: python\_repl\_ast**

**Action Input: df.groupby('Company Name').count()['ID'].sort\_values(ascending=False).head(10)**

Observation: **Company Name**

**The University Of Auckland 199**

**Fisher and Paykel Healthcare 154**

**Beca 48**

**Auckland Transport 43**

**Auckland Bioengineering Institute 36**

**Downer 29**

**Downer Nz 25**

**Aurecon 24**

**Tonkin Taylor 22**

**Wsp 19**

**Name: ID, dtype: int64**

Thought: **I now know the final answer**

**Final Answer: The top 10 companies with the most projects are The University Of Auckland, Fisher and Paykel Healthcare, Beca, Auckland Transport, Auckland Bioengineering Institute, Downer, Downer Nz, Aurecon, Tonkin Taylor, and Wsp.**

**> Finished chain.**

Out[25]:

'The top 10 companies with the most projects are The University Of Auckland, Fisher and Paykel Healthcare, Beca, Auckland Transport, Auckland Bioengineering Institute, Downer, Downer Nz, Aurecon, Tonkin Taylor, and Wsp.'