```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

from google.colab import files
uploaded = files.upload()
```

```
<IPython.core.display.HTML object>
```

```
Saving AI_Impact_on_Jobs_2030.csv to AI_Impact_on_Jobs_2030.csv
```

```python
df = pd.read_csv('AI_Impact_on_Jobs_2030.csv')

df.head()
```


```
{"summary":"{\n  \"name\": \"df\",\n  \"rows\": 3000,\n  \"fields\":
[\n    {\n      \"column\": \"Job_Title\",\n      \"properties\": {\n
\"dtype\": \"category\",\n        \"num_unique_values\": 20,\n
\"samples\": [\n          \"Security Guard\",\n          \"Chef\",\n
\"Doctor\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Average_Salary\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 34608,\n        \"min\": 30030,\n
\"max\": 149798,\n        \"num_unique_values\": 2960,\n
\"samples\": [\n          127557,\n          96783,\n          138139\
n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Years_Experience\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 8,\n        \"min\": 0,\n
\"max\": 29,\n        \"num_unique_values\": 30,\n        \"samples\":
[\n          26,\n          25,\n          3\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"Education_Level\",\n
\"properties\": {\n        \"dtype\": \"category\",\n
\"num_unique_values\": 4,\n        \"samples\": [\n          \"PhD\",\
n          \"Bachelor's\",\n          \"Master's\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"AI_Exposure_Index\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
0.28400447044949895,\n        \"min\": 0.0,\n        \"max\": 1.0,\n
\"num_unique_values\": 101,\n        \"samples\": [\n          0.14,\n
0.59,\n          0.28\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Tech_Growth_Factor\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 0.2876685590319307,\n        \"min\":
0.5,\n        \"max\": 1.5,\n        \"num_unique_values\": 101,\n
\"samples\": [\n          1.31,\n          1.39,\n          0.9\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"Automation_Probability_2030\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
```

```
0.2478812653189437,\n        \"min\": 0.05,\n        \"max\": 0.95,\n
\"num_unique_values\": 91,\n        \"samples\": [\n          0.9,\n
0.54,\n          0.34\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Risk_Category\",\n      \"properties\": {\n        \"dtype\":
\"category\",\n        \"num_unique_values\": 3,\n        \"samples\":
[\n          \"High\",\n          \"Low\",\n          \"Medium\"\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"Skill_1\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
0.2878882870909661,\n        \"min\": 0.0,\n        \"max\": 1.0,\n
\"num_unique_values\": 101,\n        \"samples\": [\n          0.46,\n
0.49,\n          0.54\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Skill_2\",\n      \"properties\": {\n        \"dtype\": \"number\",\
n        \"std\": 0.288085421288331,\n        \"min\": 0.0,\n
\"max\": 1.0,\n        \"num_unique_values\": 101,\n
\"samples\": [\n          0.39,\n          0.98,\n          0.29\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"Skill_3\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
0.28835391069799493,\n        \"min\": 0.0,\n        \"max\": 1.0,\n
\"num_unique_values\": 101,\n        \"samples\": [\n          0.15,\n
0.66,\n          0.45\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Skill_4\",\n      \"properties\": {\n        \"dtype\": \"number\",\
n        \"std\": 0.287062752541765,\n        \"min\": 0.0,\n
\"max\": 1.0,\n        \"num_unique_values\": 101,\n
\"samples\": [\n          0.5,\n          0.07,\n          0.54\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"Skill_5\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
0.28581764176034896,\n        \"min\": 0.0,\n        \"max\": 1.0,\n
\"num_unique_values\": 101,\n        \"samples\": [\n          0.3,\n
0.87,\n          0.43\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Skill_6\",\n      \"properties\": {\n        \"dtype\": \"number\",\
n        \"std\": 0.2860500611805388,\n        \"min\": 0.0,\n
\"max\": 1.0,\n        \"num_unique_values\": 101,\n
\"samples\": [\n          0.44,\n          0.97,\n          0.38\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"Skill_7\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
0.28804377700705686,\n        \"min\": 0.0,\n        \"max\": 1.0,\n
\"num_unique_values\": 101,\n        \"samples\": [\n          0.75,\n
0.87,\n          0.97\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Skill_8\",\n      \"properties\": {\n        \"dtype\": \"number\",\
n        \"std\": 0.289832048482677,\n        \"min\": 0.0,\n
```

```
\"max\": 1.0,\n        \"num_unique_values\": 101,\n
\"samples\": [\n          0.31,\n          0.43,\n          0.0\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"Skill_9\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
0.28581819904126127,\n        \"min\": 0.0,\n        \"max\": 1.0,\n
\"num_unique_values\": 101,\n        \"samples\": [\n          0.53,\n
0.42,\n          0.82\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Skill_10\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 0.286464077756422,\n        \"min\":
0.0,\n        \"max\": 1.0,\n        \"num_unique_values\": 101,\n
\"samples\": [\n          0.3,\n          0.29,\n          0.44\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    }\n  ]\n}","type":"dataframe","variable_name":"df"}
```


Propose thresholds to redefine risk categories using data.

```
riskmap = {
    'High' : 3,
    'Medium' : 2 ,
    'Low' : 1
}

df['Risk_Category'] = df['Risk_Category'].map(riskmap)

#Categorical features are converted to numerical values of Risk
Category

df.head()
```


```
{"summary":"{\n  \"name\": \"df\",\n  \"rows\": 3000,\n  \"fields\":
[\n    {\n      \"column\": \"Job_Title\",\n      \"properties\": {\n
\"dtype\": \"category\",\n        \"num_unique_values\": 20,\n
\"samples\": [\n          \"Security Guard\",\n          \"Chef\",\n
\"Doctor\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Average_Salary\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 34608,\n        \"min\": 30030,\n
\"max\": 149798,\n        \"num_unique_values\": 2960,\n
\"samples\": [\n          127557,\n          96783,\n          138139\
n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Years_Experience\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 8,\n        \"min\": 0,\n
\"max\": 29,\n        \"num_unique_values\": 30,\n        \"samples\":
[\n          26,\n          25,\n          3\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"Education_Level\",\n
\"properties\": {\n        \"dtype\": \"category\",\n
```

```
\"num_unique_values\": 4,\n        \"samples\": [\n          \"PhD\",\
n          \"Bachelor's\",\n          \"Master's\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"AI_Exposure_Index\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
0.28400447044949895,\n        \"min\": 0.0,\n        \"max\": 1.0,\n
\"num_unique_values\": 101,\n        \"samples\": [\n          0.14,\n
0.59,\n          0.28\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Tech_Growth_Factor\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 0.2876685590319307,\n        \"min\":
0.5,\n        \"max\": 1.5,\n        \"num_unique_values\": 101,\n
\"samples\": [\n          1.31,\n          1.39,\n          0.9\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"Automation_Probability_2030\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
0.2478812653189437,\n        \"min\": 0.05,\n        \"max\": 0.95,\n
\"num_unique_values\": 91,\n        \"samples\": [\n          0.9,\n
0.54,\n          0.34\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Risk_Category\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 0,\n        \"min\": 1,\n
\"max\": 3,\n        \"num_unique_values\": 3,\n        \"samples\":
[\n          3,\n          1,\n          2\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"Skill_1\",\n      \"properties\":
{\n        \"dtype\": \"number\",\n        \"std\":
0.2878882870909661,\n        \"min\": 0.0,\n        \"max\": 1.0,\n
\"num_unique_values\": 101,\n        \"samples\": [\n          0.46,\n
0.49,\n          0.54\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Skill_2\",\n      \"properties\": {\n        \"dtype\": \"number\",\
n        \"std\": 0.288085421288331,\n        \"min\": 0.0,\n
\"max\": 1.0,\n        \"num_unique_values\": 101,\n
\"samples\": [\n          0.39,\n          0.98,\n          0.29\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"Skill_3\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
0.28835391069799493,\n        \"min\": 0.0,\n        \"max\": 1.0,\n
\"num_unique_values\": 101,\n        \"samples\": [\n          0.15,\n
0.66,\n          0.45\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Skill_4\",\n      \"properties\": {\n        \"dtype\": \"number\",\
n        \"std\": 0.287062752541765,\n        \"min\": 0.0,\n
\"max\": 1.0,\n        \"num_unique_values\": 101,\n
\"samples\": [\n          0.5,\n          0.07,\n          0.54\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"Skill_5\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
```

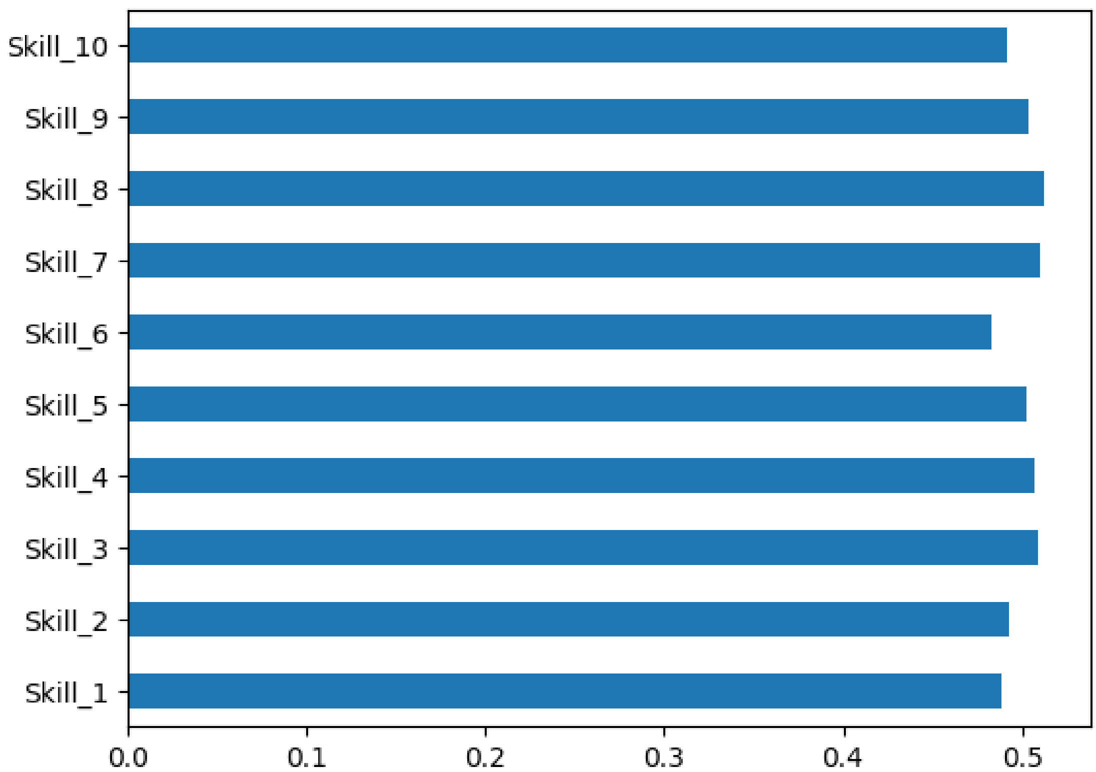
```
0.28581764176034896,\n        \"min\": 0.0,\n        \"max\": 1.0,\n
\"num_unique_values\": 101,\n        \"samples\": [\n          0.3,\n
0.87,\n          0.43\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Skill_6\",\n      \"properties\": {\n        \"dtype\": \"number\",\
n        \"std\": 0.2860500611805388,\n        \"min\": 0.0,\n
\"max\": 1.0,\n        \"num_unique_values\": 101,\n
\"samples\": [\n          0.44,\n          0.97,\n          0.38\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"Skill_7\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
0.28804377700705686,\n        \"min\": 0.0,\n        \"max\": 1.0,\n
\"num_unique_values\": 101,\n        \"samples\": [\n          0.75,\n
0.87,\n          0.97\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Skill_8\",\n      \"properties\": {\n        \"dtype\": \"number\",\
n        \"std\": 0.289832048482677,\n        \"min\": 0.0,\n
\"max\": 1.0,\n        \"num_unique_values\": 101,\n
\"samples\": [\n          0.31,\n          0.43,\n          0.0\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"Skill_9\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
0.28581819904126127,\n        \"min\": 0.0,\n        \"max\": 1.0,\n
\"num_unique_values\": 101,\n        \"samples\": [\n          0.53,\n
0.42,\n          0.82\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Skill_10\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 0.286464077756422,\n        \"min\":
0.0,\n        \"max\": 1.0,\n        \"num_unique_values\": 101,\n
\"samples\": [\n          0.3,\n          0.29,\n          0.44\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    }\n  ]\n}","type":"dataframe","variable_name":"df"}
```


Which skills (Skill_1 to Skill_10) are most associated with low automation risk?

```
gp = df.groupby("Risk_Category").agg({"Skill_1" : 'mean' ,
"Skill_2" : 'mean' , "Skill_3" : 'mean' , 'Skill_4' : 'mean' ,
'Skill_5':'mean' , 'Skill_6':'mean' , 'Skill_7' : 'mean' , 'Skill_8':
'mean' , 'Skill_9' : 'mean' , 'Skill_10' : 'mean'}).loc[1]

gp.plot(kind = 'barh')
```

```
<Axes: >
```




```
#Skills 8, 7, and 3 show the strongest association with low automation
risk, while Skills 6 and 1 are comparatively more automatable.
```

Identify jobs that meet all three criteria:

High salary

Low automation probability

High tech growth factor

```
high_s = df['Average_Salary'].quantile(0.75)
low_at = df['Automation_Probability_2030'].quantile(0.25)
high_tech = df['Tech_Growth_Factor'].quantile(0.75)

filtered = df[
    (df['Average_Salary'] >= high_s ) &
    (df['Automation_Probability_2030'] <= low_at) &
    (df['Tech_Growth_Factor'] >= high_tech)
][['Job_Title', 'Average_Salary', 'Automation_Probability_2030',
'Tech_Growth_Factor']]

filtered['Job_Title'].unique()

array(['Doctor', 'Research Scientist', 'Lawyer', 'Teacher', 'Nurse',
       'AI Engineer', 'Mechanic'], dtype=object)
```

```
# The Doctor', 'Research Scientist', 'Lawyer', 'Teacher', 'Nurse','AI
Engineer', 'Mechanic' are the jobs match criteria
```