The Job Explorer
— A $alary Analysis

Group Bambi : Maggie Xu, Lia Wang, Ceci Chen





Table of Contents

Data Preprocessing and Cleaning

Dataset Salary_data_based_on_country_and_race(2023)

	Unnamed: 0	Age	Gender	Education Level	Job Title	Years of Experience	Salary	Country	Race
0	0	32	Male	Bachelor's	Software Engineer	5.0	$90,000.00	UK	White
1	1	28	Female	Master's	Data Analyst	3.0	$65,000.00	USA	Hispanic
2	2	45	Male	PhD	Senior Manager	15.0	$150,000.00	Canada	White
3	3	36	Female	Bachelor's	Sales Associate	7.0	$60,000.00	USA	Hispanic
4	4	52	Male	Master's	Director	20.0	$200,000.00	USA	Asian

Source: Hosted on Kaggle, originally compiled by VIRA SPASIVSKA

Overview: A comprehensive collection of salary information from various industries and regions across the globe

Dimension: 6700 rows * 9 columns

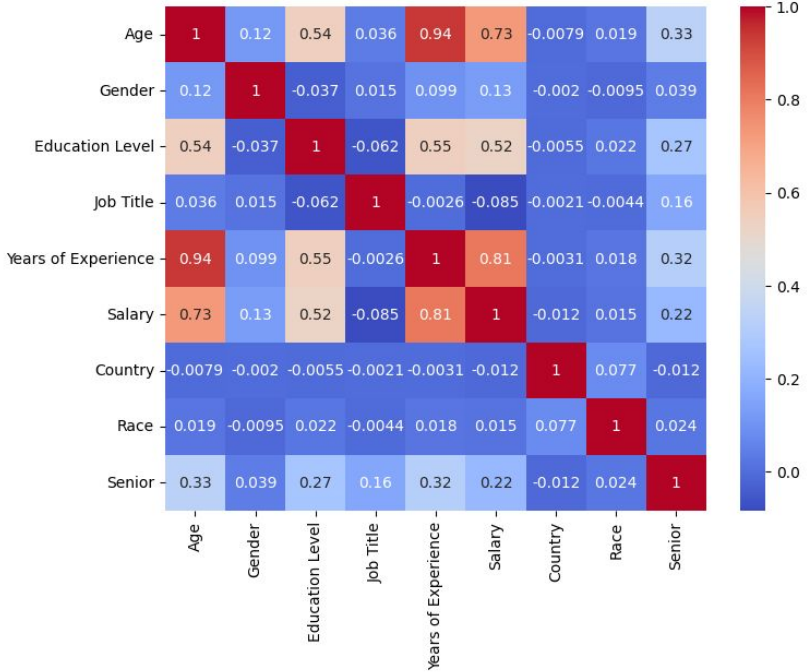
Cleaning

- Drop the index column (Unnamed)
- Drop the missing values (only 2 in Salary)
- Convert Salary data to float
- Filter out Gender = Other
- Add a dummy variable 'Senior' to identify Senior titles

Column	Description	Type
Age	Employee's Age (21 - 62)	int64
Gender	Female or Male	object
Education Level	High School, Bachelor's, Master's, PHD	object
Job Title	129 Different Job Titles	object
Years of Experience	Number of years of working Experience (0 - 34)	float64
Salary	Annual salary (350 - 250000)	float64
Country	USA(1356), China(1339), Canada(1322), UK(1332), Australia(1335)	object
Race	White, Hispanic, Asian, Black, Others (Welsh, Mixed, Australian)	object
Senior	Whether the job is a senior title (== 1) or not (== 0)	int64

Research Question 1

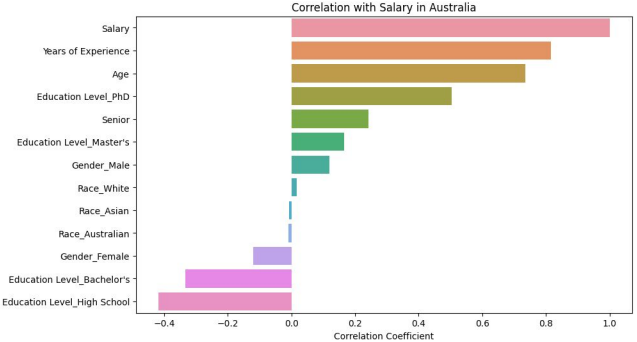
Generally, how does each feature correlate with each other?

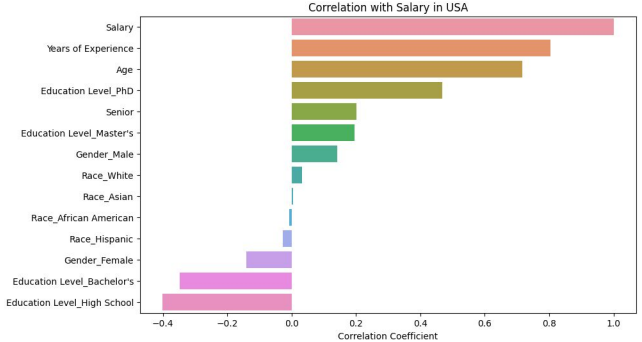


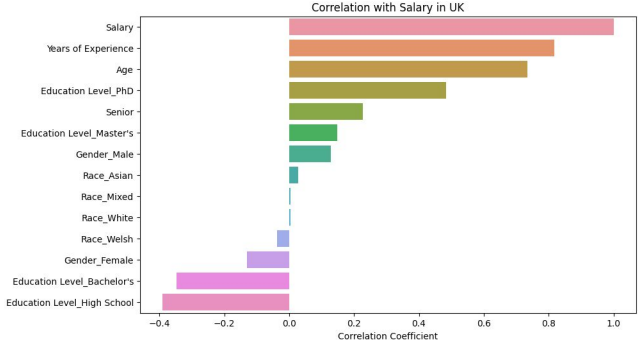
Remarks

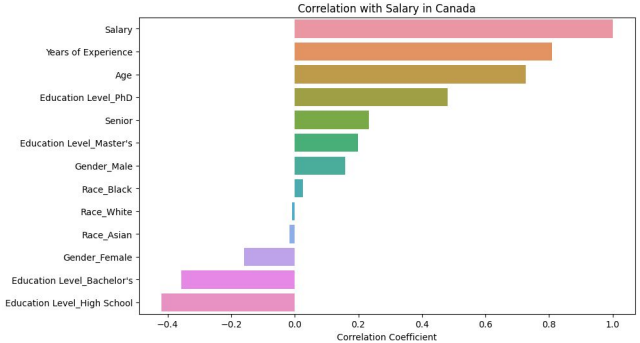
Age & Years of Experience (0.94)
Salary & Years of Experience (0.81)
Age & Salary (0.73)
Years of Experience & Education Level (0.55)
Age & Education Level (0.54)
Salary & Education Level (0.52)

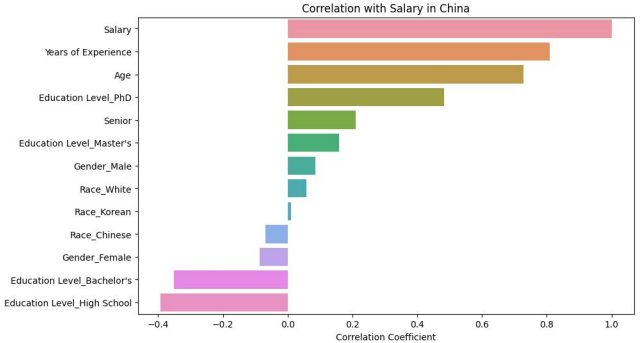
Country Comparisons





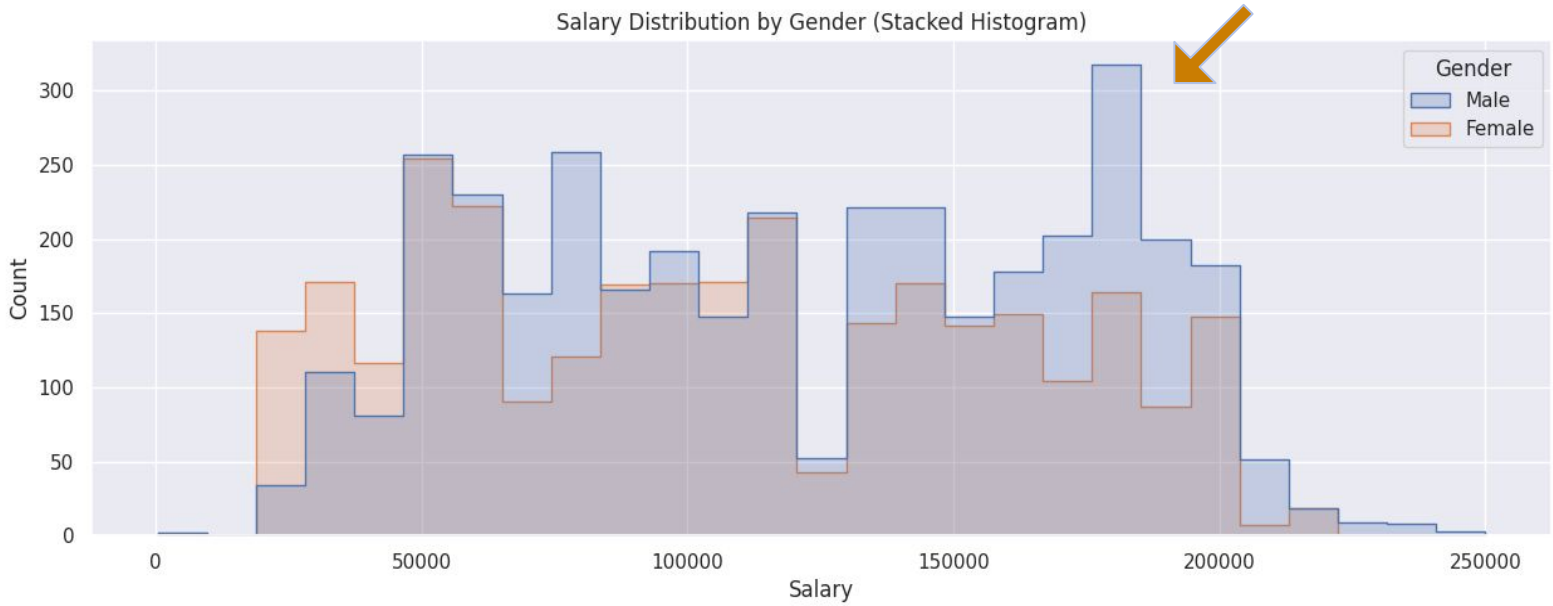




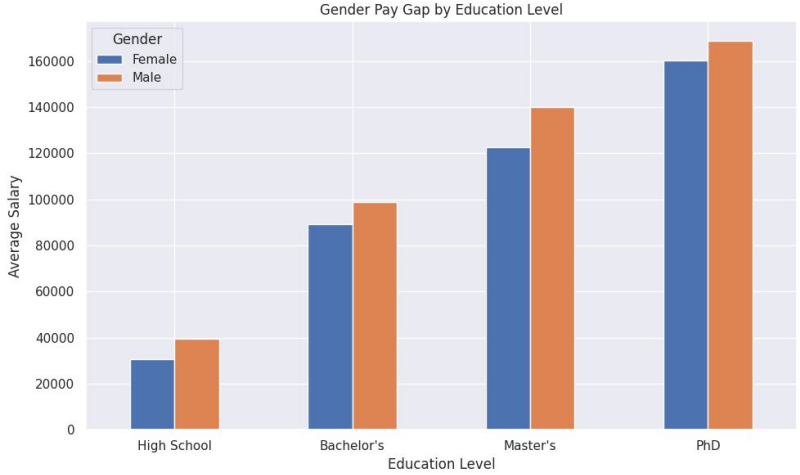


Research Question 2

How are salaries distributed among males and females? Does the *Gender Pay Gap* exist?



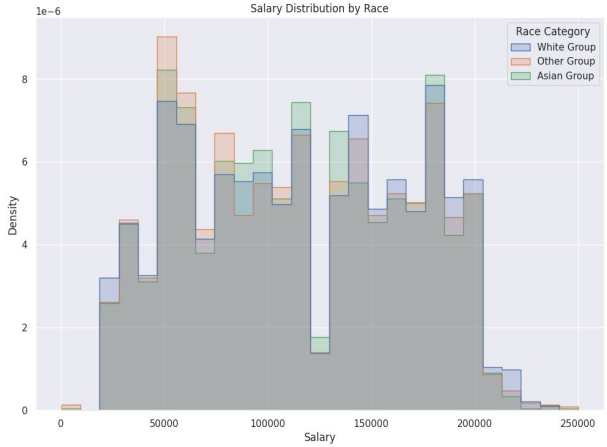
Is there a *Gender Pay Gap* across various education levels and years of experiences?

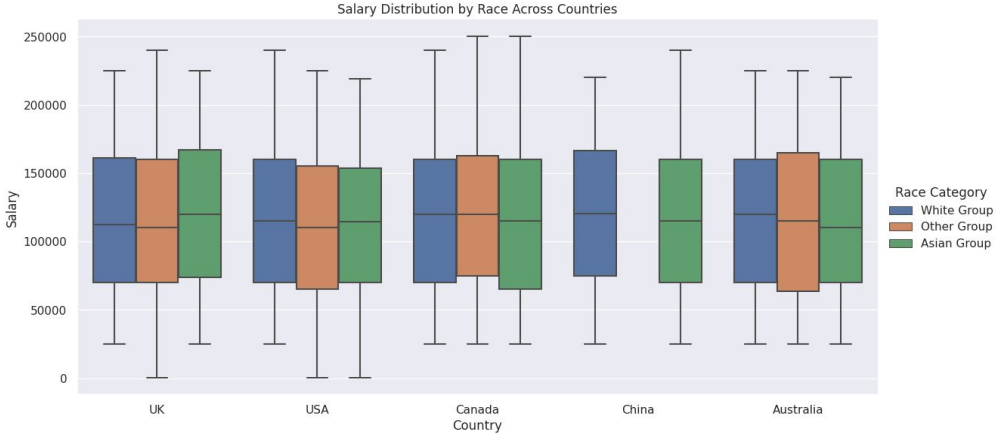


On average, males earn more than females, and this difference is present across various education levels and years of experience.

Research Question 3

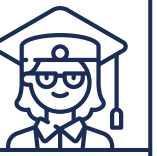
Group by *Asian, White, and Other Races*; How are salaries distributed among different racial groups? In which country, Asians (since we identify ourselves as Asians) earn the highest salary?





Asian relatively have higher concentration of individuals earning in middle salary range(75,000 - 14,0000). On average, Asians earn the highest salary in UK.

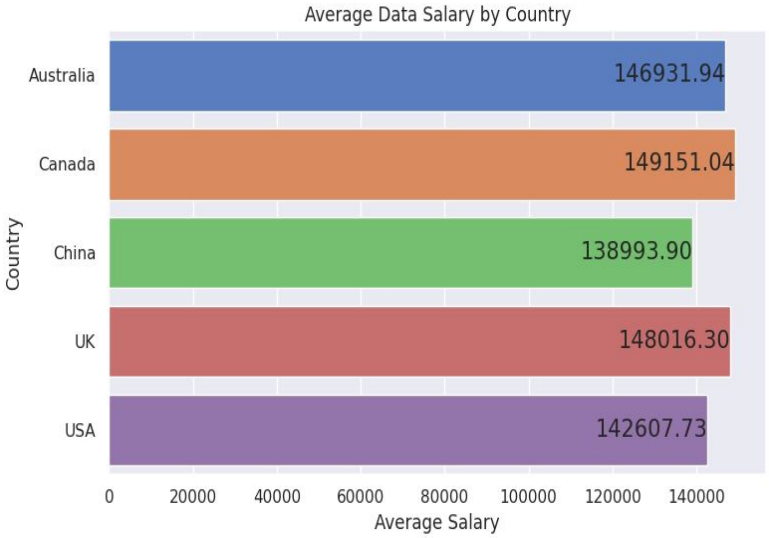







Research Question 4

For jobs in Data Industry (*Data Analyst, Data Engineer, Data Entry Clerk, Data Scientist, Chief Data Officer*), which country (China or USA) has relative higher salary?

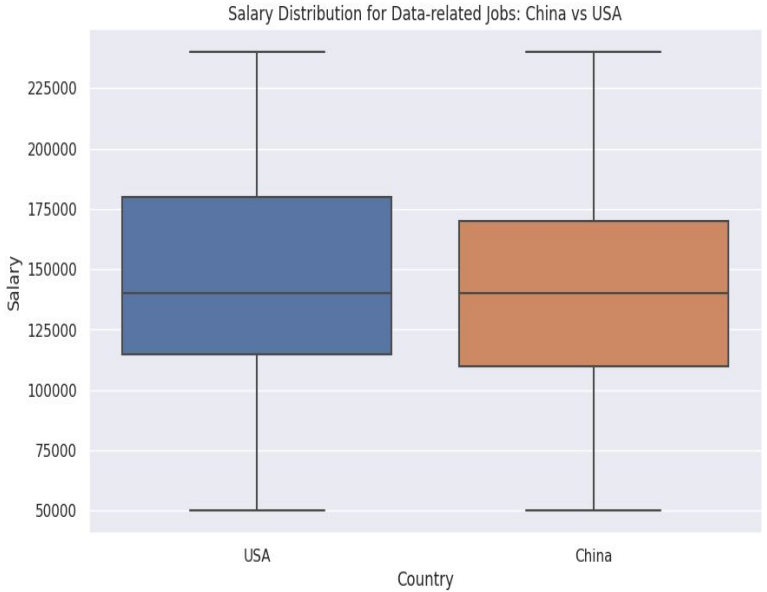


👈 Average Data Salary by Country

- **Average Data Salary in China & USA are relatively lower than the other countries**

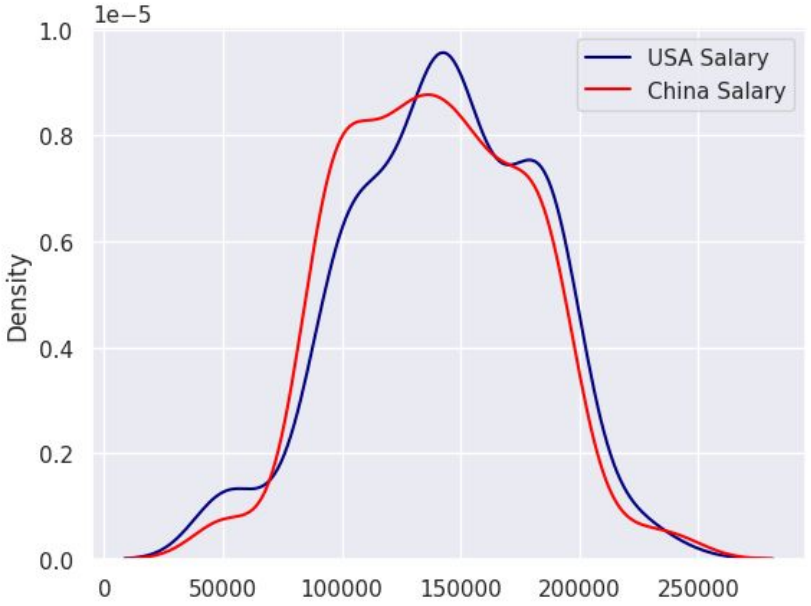
Salary Distribution: China vs USA👉

- **No obvious pay gap between two countries**



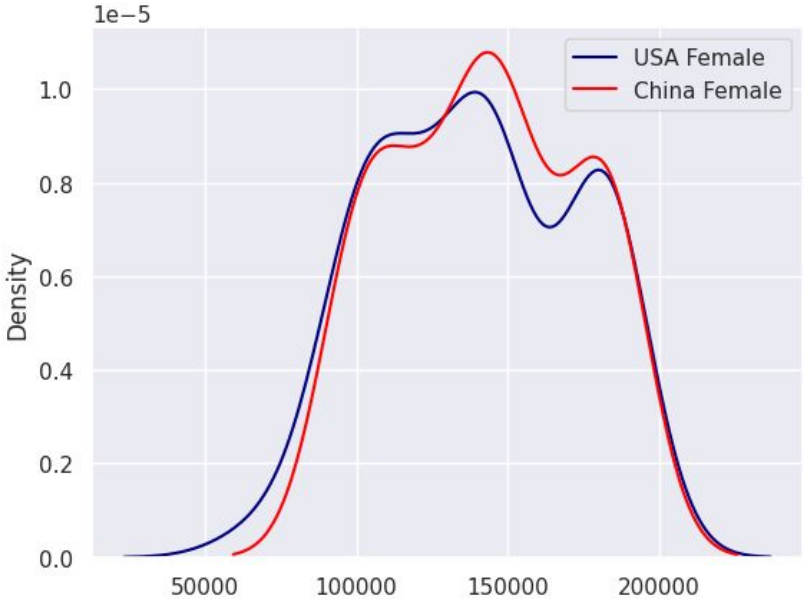
Hypothesis Test 1: Is there a significant difference on Data Industry payment in China & USA?

- **H0**: There is no significant difference in Data industry payment in USA and China.
- **H1**: Data industry jobs earn more in USA than China
- P-value = 0.11597032 > 0.05
- Fail to reject the null hypothesis.
- **Conclusion:** There is no significant difference of Data industry payment in USA and China.



Hypothesis Test 2: Is there a significant difference on Data Industry payment for female in China & USA?

- **H0**: There is no significant difference in Data industry payment for female in USA and China.
- **H1**: Female in Data industry earn more in USA than China
- P-value = 0.71771199 > 0.05
- Fail to reject the null hypothesis.
- **Conclusion:** There is no significant difference in Data industry payment for female in USA and China.



Machine Learning Models For Prediction

- Predictors (X): ***'Race', 'Country', 'Education Level', 'Gender', 'Job Title'***
  - Exclude strongly positive correlated features: *'Age', 'Years of Experience', 'Senior'*
  - Exclude duplicate feature: *'Race Category'*
- Target (Y): ***'Salary'***
- 80% Train Data 20% Test Data
- Models: **Linear Regression Model** & **Random Forest Model**
- Data Used to Predict:
  - ***pairwise('Race', 'Country') * 'Job Title' * 'Gender' * 'Education Level'***
  - Shape: (17544, 5)

	Race	Country	Education Level	Gender	Job Title
0	Australian	Australia	Bachelor's	Male	Software Engineer
1	Australian	Australia	Bachelor's	Male	Data Analyst
2	Australian	Australia	Bachelor's	Male	Manager
3	Australian	Australia	Bachelor's	Male	Sales Associate
4	Australian	Australia	Bachelor's	Male	Director

Linear Regression

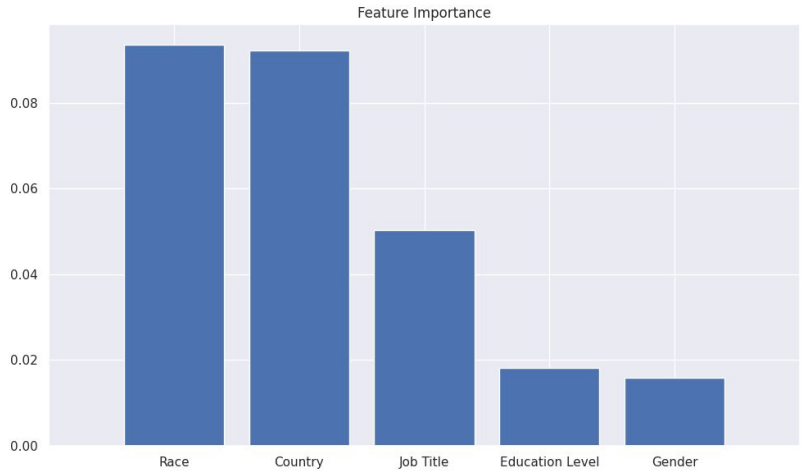
- R-squared: 0.638

- Intercept: 98568.701
- Coefficients:
  - Gender: -1915.106
  - Education Level: 1915.106
  - Job Title: -7739.664
  - Country: -35496.737
  - Race: 14747.958

Random Forest

- R-squared: 0.639

- Feature Importances
  - Gender: 0.0169
  - Education Level: 0.0173
  - Job Title: 0.0409
  - Country: 0.0921
  - Race: 0.1020



Conclusion of Prediction:

- Choose to use **Random Forest Model (2)**
- **Asian, USA, Phd, Male, Chief Technology Officer**
- Predicted Salary: **$218,955**

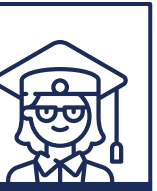



🏆100%🏆

Further Discussion

- **Model Improvement:** Tuning hyperparameter to improve model performance or try more models.
- **Industry Trends**: Every field is rapidly evolving, and industry demand can significantly influence salary.
- **Specific Skills**: More specific data, such as skills, could offer more relevant insights.
- **Work Authorization**: For international job search, work visa regulations and the ease of obtaining work authorization should be taken into account
- **Anti-discrimination Laws and Policies**: Legal framework regarding employment discrimination in different countries can be important.





Thank you