FINAL PRESENTATION MARCH 25, 2024

Group 3: Christopher Murphy, Daylin Hernandez, Christopher Cavino and Chad Alessi

SKILLS MIGRATION TRENDS OF LINKEDIN MEMBER PROFILES (2015-2019)

- Introduction
- Business Questions
- Data Set Retrieval & Data Munging
- Descriptive Statistics & Visualizations
- Data Modeling & Visualizations
- Summary

INTRODUCTION

- This project analyzed the patterns of skills migration of LinkedIn member data over five years.
- The years included from 2015 to 2019.
- We aimed to uncover the underlying trends of professional mobility on a global scale.
- The movements of skilled labor, categorized by various attributes, including country code, country name, World Bank income classification, World Bank region, and specific skill groups.

INITIAL BUSINESS QUESTIONS

- What are the key attributes?
- What are the trends?
- What are the top N industries experiencing the highest change?
- Which region was most impacted by migration trends?
- What are the highest growth rates for the top 10 careers?
- How do the labor and skill loss data align to the industry growth trends?

DATA SET RETRIEVAL & DATA MUNGING

- Read in the excel dataset and saved it in a data frame called skillMigrationDF.
- We Checked for any missing information or NAs in the dataset using colSums(is.na()).
    We kept all data from the original dataset. No data was replaced or removed.
- Separated out the individual regions.
    East Asia & Pacific - skillEAP.df
    Europe & Central Asia - skillECA.df
    Latin America & Caribbean - skillLAC.df
    Middle East & North Africa - skillMENA.df
    North America - skillNA.df
    South Asia - skillSA.df
    Sub-Saharan Africa - skillSSA.df
- Created a new column called net_per_10k_cumulative. This is the sum of net_per_10k_2015, net_per_10k_2016, net_per_10k_2017, net_per_10k_2018, and net_per_10k_2019.
- Create a new column called income_class. Where 0 means that the observation is classified under lower income or lower middle income. If 1, the observation is classified as high or upper middle income.
- Create another column net_per_10k_2019_binary. A 0 means a negative net_per_10k_2019, and a 1 is a positive net_per_10k_2019.

DATA SET RETRIEVAL & DATA MUNGING

Skill Migration	
country_code & *country_name*	Country name given by World Bank taxonomy, and 2 letter country code
wb_region & *wb_income*	(World Bank Region & World Bank Income Group) Country categories classified by the latest World Bank region and income group
skill_group_name	Skill groups categorize the 50,000 detailed individual skills into approximately 250 skills groups (skill groups may be excluded based data quality considerations). For example, web development (skill group), may be composed of java, html etc. skills).
net_per_10K_YYYY	Absolute '*netflow_YYYY*' divided by '*total_member_ct_YYYY*', with respect to '*country_name*' and '*skill_group_name*', for specified year. Rate multiplied by 10,000 to simplify interpretation.

```
```{r}
str(skillMigrationDF)
```
```

```
'data.frame':   17617 obs. of  12 variables:
 $ country_code        : chr  "af" "af" "af" "af" ...
 $ country_name        : chr  "Afghanistan" "Afghanistan" "Afghanistan" "Afghanistan" ...
 $ wb_income           : chr  "Low income" "Low income" "Low income" "Low income" ...
 $ wb_region           : chr  "South Asia" "South Asia" "South Asia" "South Asia" ...
 $ skill_group_id      : num  2549 2608 3806 50321 1606 ...
 $ skill_group_category: chr  "Tech Skills" "Business Skills" "Specialized Industry Skills" "Tech Skills" ...
 $ skill_group_name    : chr  "Information Management" "Operational Efficiency" "National Security" "Software Testing"
...
 $ net_per_10K_2015    : num  -792 -1610 -1731 -958 -1511 ...
 $ net_per_10K_2016    : num  -706 -934 -770 -829 -841 ...
 $ net_per_10K_2017    : num  -550 -776 -757 -965 -842 ...
 $ net_per_10K_2018    : num  -681 -532 -600 -406 -582 ...
 $ net_per_10K_2019    : num  -1209 -790 -768 -740 -719 ...
```

DESCRIPTIVE STATISTICS & VISUALIZATIONS

Average skill migration per year in each region
(in order as followed: 2015,2016,2017,2018,2019)

averageEAP	averageECA	averageLAC	averageMENA	averageNA	averageSA	averageSSA
43.28502	-15.4445564	-74.59658	6.533663	17.26113	-216.2011	32.35592
25.14751	-0.8243587	-121.26805	-75.543073	33.38045	-206.3829	-44.81603
-15.71553	9.9298714	-146.41812	-118.054603	43.98405	-181.8076	-82.27412
25.14251	33.8352893	-182.72687	-77.502287	56.51573	-151.9598	-75.29864
14.52680	44.9459627	-172.22399	-85.639665	65.59516	-189.5543	-71.27973

DESCRIPTIVE STATISTICS & VISUALIZATIONS

Country with the highest migration each year from every region

country_code	country_name	wb_income	wb_region	skill_group_id	skill_group_category	skill_group_name	net_per_10K_2015	net_per_10K_2016	net_per_10K_2017	net_per_10K_2018	net_per_10K_2019
ml	Mali	Low income	Sub-Saharan Africa	1655	Specialized Industry Skills	Army	2824.97	-1479.05	1906.14	76.53	-732.60
lu	Luxembourg	High income	Europe & Central Asia	921	Soft Skills	Teamwork	1657.96	1796.89	1572.35	1433.05	1345.65
ml	Mali	Low income	Sub-Saharan Africa	1655	Specialized Industry Skills	Army	2824.97	-1479.05	1906.14	76.53	-732.60
lu	Luxembourg	High income	Europe & Central Asia	20581	Specialized Industry Skills	Analytical Reasoning	1125.18	1646.73	1069.69	1515.79	1311.73
ge	Georgia	Lower middle income	Europe & Central Asia	2591	Business Skills	Customer Experience	236.22	-38.81	-205.13	781.72	1901.99

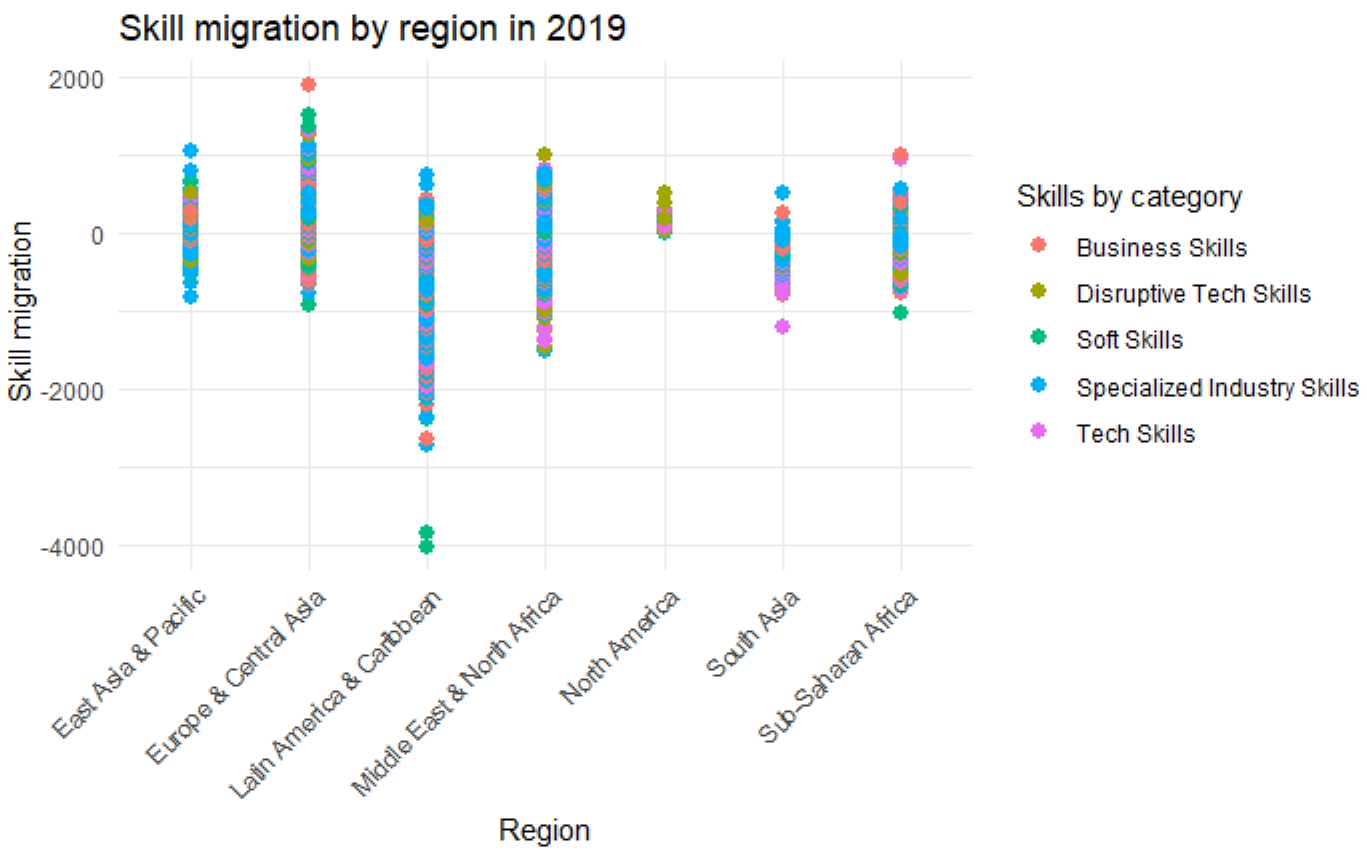
Europe and Central Asia had the highest skill migration in three of the five years.
Specifically, Mali was a country that had the maximum skill migration in two of the five years.
Specialized Industry skills was the group category listed most frequently.

Country with the lowest migration each year from every region

country_code	country_name	wb_income	wb_region	skill_group_id	skill_group_category	skill_group_name	net_per_10K_2015	net_per_10K_2016	net_per_10K_2017	net_per_10K_2018	net_per_10K_2019
cu	Cuba	Upper middle income	Latin America & Caribbean	50304	Tech Skills	Mobile Application Development	-3037.38	-1968.18	-1155.10	-1332.28	-1708.36
ve	Venezuela, RB	Upper middle income	Latin America & Caribbean	31090	Soft Skills	Flexible Approach	-1209.32	-2435.26	-2542.23	-3629.02	-4022.04
dz	Algeria	Upper middle income	Middle East & North Africa	44	Specialized Industry Skills	Recruiting	-55.45	-100.74	-6604.67	-151.61	-260.71
ve	Venezuela, RB	Upper middle income	Latin America & Caribbean	31090	Soft Skills	Flexible Approach	-1209.32	-2435.26	-2542.23	-3629.02	-4022.04
ve	Venezuela, RB	Upper middle income	Latin America & Caribbean	31090	Soft Skills	Flexible Approach	-1209.32	-2435.26	-2542.23	-3629.02	-4022.04

Latin America and Caribbean had the lowest value of skill migration in four of the five years.
Venezuela was repeated in three of the five years as the country experiencing more skill loss.
Soft skills such as a flexible approach were demonstrated to be the most impacted.
The income class with minimum values of skill migration was the upper middle income for the five years.

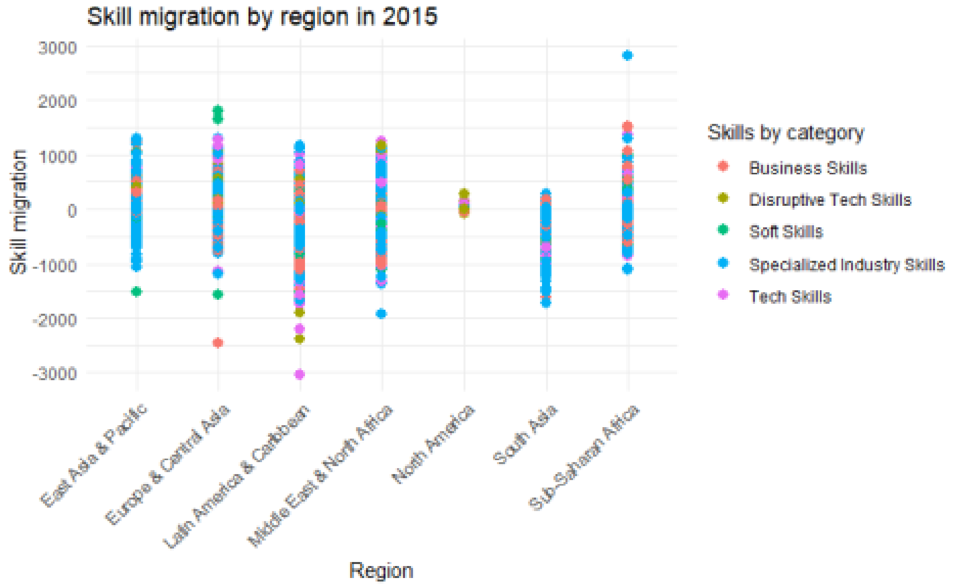
DESCRIPTIVE STATISTICS & VISUALIZATIONS

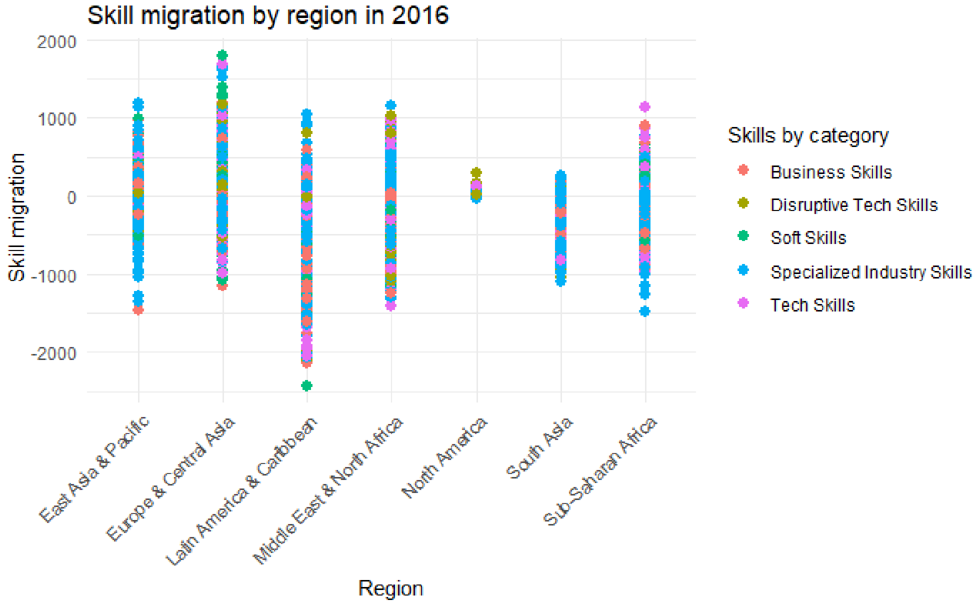


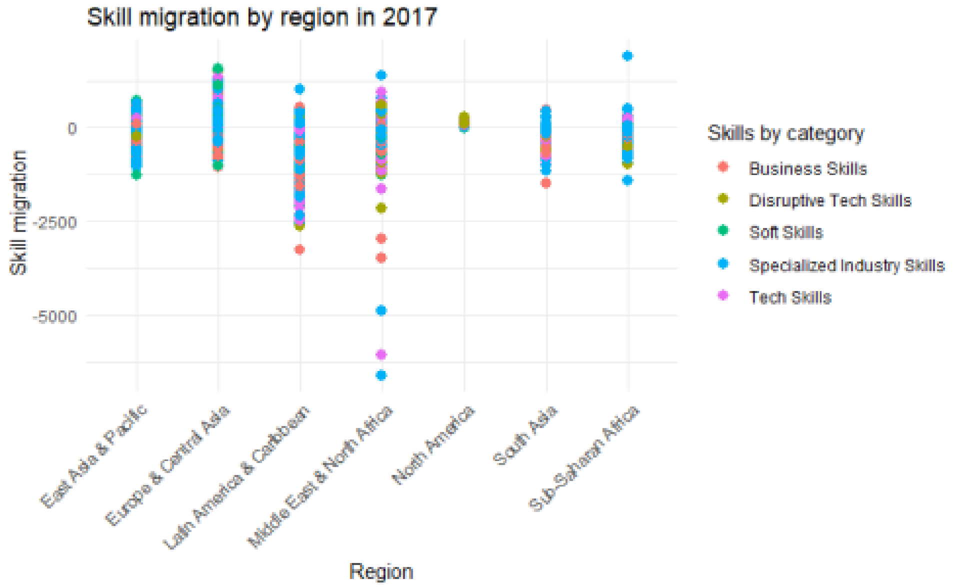
- Business Skills (ex. Human Resources, Investment Banking)
- Disruptive Tech Skills (ex. Data Science, Aerospace Engineering)
- Soft Skills (ex. Writing, Leadership)
- Specialized Industry Skills (ex. Law, Drilling Engineering)
- Techn Skills (ex. System Administration, Social Media)

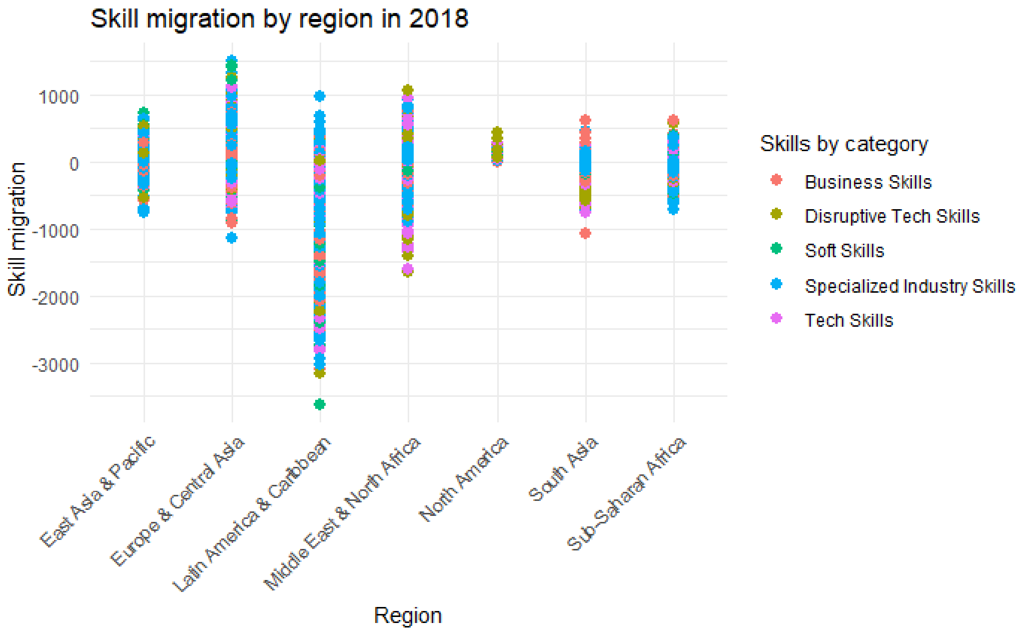
Shows the skills gained and lost during 2019. Latin America & Caribbean lost the most, whereas Europe & Central Asia gained the most.

DESCRIPTIVE STATISTICS & VISUALIZATIONS









LINEAR PREDICTION MODELING

Linear regression was successful for predicting net skills migration for the final year of the study.

Predicting net migration per 10,000 people in 2019 (net_per_10K_2019) using the net migration rates from 2015 to 2018 as predictors:

```
Call:
lm(formula = net_per_10K_2019 ~ net_per_10K_2015 + net_per_10K_2016 +
    net_per_10K_2017 + net_per_10K_2018, data = skillMigrationDF1)

Residuals:
     Min       1Q   Median       3Q      Max
-1677.78   -42.89     2.49    39.72  1382.51

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      -4.204348   0.813821  -5.166 2.42e-07 ***
net_per_10K_2015  0.088106   0.005128  17.183  < 2e-16 ***
net_per_10K_2016  0.069205   0.006868  10.077  < 2e-16 ***
net_per_10K_2017  0.050497   0.006018   8.391  < 2e-16 ***
net_per_10K_2018  0.659981   0.005813 113.539  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 104.7 on 17612 degrees of freedom
Multiple R-squared:  0.8095,    Adjusted R-squared:  0.8094
F-statistic: 1.871e+04 on 4 and 17612 DF,  p-value: < 2.2e-16
```

Insights:

- The ability to **predict the gain/loss of skills for a coming year** could serve as valuable input to policy makers, educators and employers attempting to respond to changes in skill migration into or out of a country.
- The linear regression equation could be useful in predicting skills trends, **providing data-driven insight** to ongoing skills changes experienced by industries.
- Predictions can be valuable input to shape **approaches to attract, develop or retain critical skills** required to enable industry

Prediction Equation:

Y = -4.20438
+0.088106*X_*2015*+0.069205*X_*2016*+0.050497*X_*2017*+0.659981*X_*2018*

SUPPORT VECTOR MACHINE CLASSIFICATION MODELING

An SVM was successful in classifying skill migration data by country income level.

Predicting "higher income" vs "lower income" country classification based upon the annual migration data (2015-2019):

```
Confusion Matrix and Statistics

          Reference
Prediction    0    1
         0 1313    2
         1    0 3969

               Accuracy : 0.9996
                 95% CI : (0.9986, 1)
    No Information Rate : 0.7515
    P-Value [Acc > NIR] : <2e-16

                  Kappa : 0.999

 Mcnemar's Test P-Value : 0.4795

            Sensitivity : 1.0000
            Specificity : 0.9995
         Pos Pred Value : 0.9985
         Neg Pred Value : 1.0000
             Prevalence : 0.2485
         Detection Rate : 0.2485
   Detection Prevalence : 0.2489
      Balanced Accuracy : 0.9997

       'Positive' Class : 0
```

Insights:

- The ability to **classify the income level of the country** generating migration data may provide insight to how income affects the gain or loss of critical skills.
- Gaining insight into **skills migration trends between countries of at the same or different income levels** may provide a valuable starting point to policy makers investigating market forces driving migration.
- Migration data for specific skills could then be analyzed to inform country strategy / policy in the face of migration trends for critical skills.

R-code:

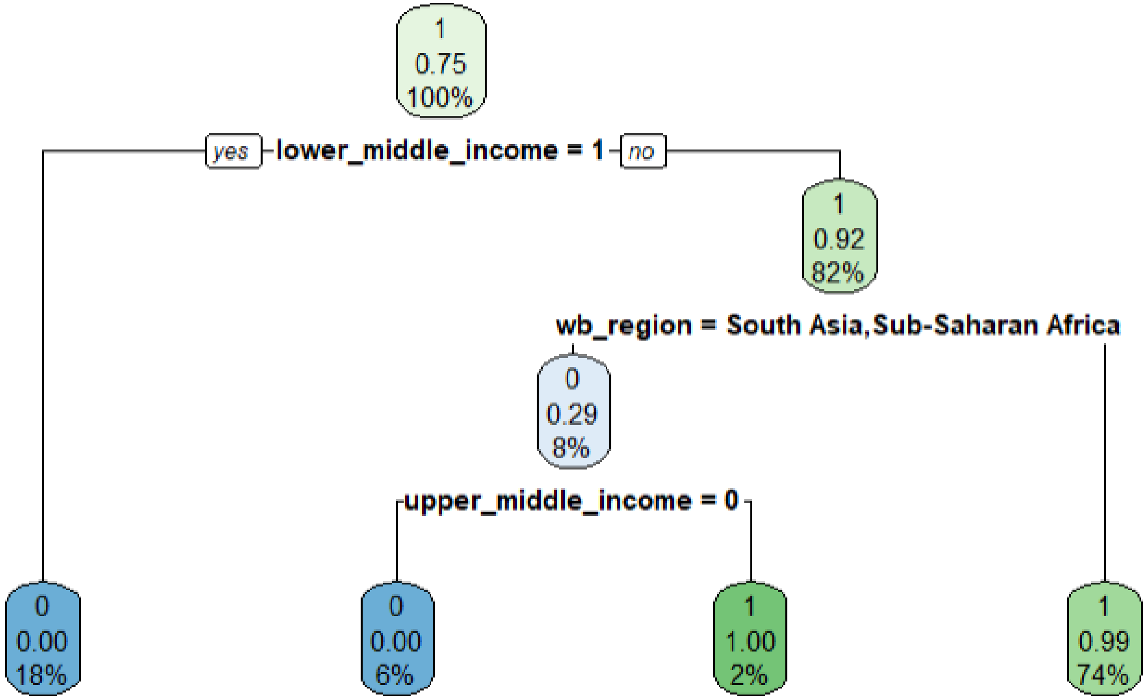
```
trainList <- createDataPartition(y=modelingDataFieldsDF$blended_higher_income,p=0.70,list=FALSE)
#train the model
svm.model.blendedHigherIncome <- train(blended_higher_income~.,
                   data=trainingDataSet,method="svmRadial",
                   trControl=trainControl(method="none"),
                   preProc=c("center","scale"))
```

RECURSIVE PARTITIONING CLASSIFICATION MODELING

Recursive partitioning was successful in classifying skill migration data by country income level.

Classifying "higher income" vs "lower income" countries based upon the annual migration data (2015-2019):

```
  predictions
     0    1
0 1266   47
1    0 3971
```



<u>Insights</u>:

- The ability to **classify the income level of the country** generating migration data could provide insight to how income affects the gain or loss of critical skills.
- Gaining insight into **skills migration trends between countries of at the same or different income levels** could provide a valuable starting point to policy makers investigating market forces driving migration.
- Migration data for specific skills could then be analyzed to inform country strategy / policy in the face of migration trends for critical skills.

R-code:

```
```{r}
#build the model
model.rpart.blendedHigherIncome <- rpart(blended_higher_income~.,data=trainingDataSet,method="class")

#visualize the results using rpart.plot()
rpart.plot(model.rpart.blendedHigherIncome)
```
```

SUMMARY

- Explored global skill migration from 2015-2019 using LinkedIn data, uncovering key insights into talent mobility.
- Found that economic status and region significantly influence skill migration, with Europe and Central Asia seeing the highest influx, and Latin America and the Caribbean the lowest.
- Revealed that areas with the least skill migration are mostly upper middle-income, with specifics on Europe, Central Asia, and Latin America.
- Employed predictive models like linear modeling and support vector machines to quantify and forecast skill migration trends effectively.
- Concluded that understanding skill migration is crucial for policymakers, businesses, and educational institutions to develop effective strategies and align with market needs.

APPENDIX

Supplemental Information

DATA SOURCE LOCATIONS

[Skill Migration ("Skill Migration" sheet)](https://datacatalogfiles.worldbank.org/ddh-published/0038044/DR0046256/public_use-talent-migration.xlsx?versionId=2024-02-13T16:57:39.2869535Z)
https://datacatalogfiles.worldbank.org/ddh-published/0038044/DR0046256/public_use-talent-migration.xlsx?versionId=2024-02-13T16:57:39.2869535Z

The team also investigated 3 additional data sets from this study. However, Skills Migration was chosen for the final project as the data set contained 17,600+ observations. Project Update 2 included analysis of the below portions of the study. That analysis allowed the team to drive to a decision for the focus of the project.

Industry Employment Growth
https://datacatalogfiles.worldbank.org/ddh-published/0038045/DR0046261/public_use-industry-employment-growth.xlsx?versionId=2023-01-19T09:28:51.4382607Z

Skills Needs
https://datacatalogfiles.worldbank.org/ddh-published/0038027/DR0046191/public_use-industry-skills-needs.xlsx?versionId=2023-01-19T03:44:21.4528001Z

Skills Penetration
https://datacatalogfiles.worldbank.org/ddh-published/0038027/DR0046192/public_use-skill-penetration.xlsx?versionId=2023-01-19T03:44:28.0600027Z

THANK YOU

Daylin Hernandez | dherna38@syr.edu

Christopher Cavino | clcavino@syr.edu

Chad Alessi | cjalessi@syr.edu

Christopher Murphy | cmurph66@syr.edu