
The Best Location
For A New
Opera
Company

The premise:

Where are the best existing and emerging markets for Opera to create a new traveling Opera Company that takes advantage of existing patterns in these markets?

This study is designed to engage and inform opera lovers, regardless of their level of experience, making it accessible to a wider audience interested in the art form's growth and future.

The recommendation is to create a traveling Opera Company:

1. Use well-established opera hubs during “peak season”:
   - Italy and Austria
2. Take the same productions to emerging markets during “off-peak seasons”:
   - Sweden, the Netherlands, Poland, Estonia, and Latvia

How the data started:

	index	season	iso	city	composer	db	dd	nat	mf	work	worknat	type	start date	performances	production
0	0	1213	al	Tirana	Lortzing	18011023	18510121	de	m	Ali Pascha von Janina	de	~B	20130323	4	NaN
1	1	1213	al	Tirana	Mozart	17560127	17911205	at	m	Don Giovanni	it	~	20130518	3	NaN
2	2	1213	al	Tirana	Puccini	18581222	19241129	it	m	Tosca	it	~	20130213	2	NaN
3	3	1213	am	Yerevan	Spendiaryan	18711020	19280507	am	m	Almast	am	~	20130711	1	NaN
4	4	1213	am	Yerevan	Tigranian	18791226	19500210	am	m	Anoush	am	~	20130511	3	NaN

How the data ended:

season	iso	city	work	start date	erformance	ountry Nam	ty populati	ntry popula	continent	sub-region	es_season_	r_10k_ppl_	nces_seaso	r_1k_ppl_c	a_by_comp	season_by	es_season_	al_per_1k_	al_per_10k	Season Year
1213	al	Tirana	Ali Pascha v	##########	4	Albania	418495	2854907	Europe	Southern E	9	0.031525	9	0.021506	Ali Pascha v	110	110	0.262847	0.385301	##########
1213	al	Tirana	Don Giovar	##########	3	Albania	418495	2854907	Europe	Southern E	9	0.031525	9	0.021506	Don Giovar	110	110	0.262847	0.385301	##########
1213	al	Tirana	Tosca	##########	2	Albania	418495	2854907	Europe	Southern E	9	0.031525	9	0.021506	Tosca by Pu	110	110	0.262847	0.385301	##########
1213	am	Yerevan	Almast	##########	1	Armenia	1093485	2990731	Asia	Western As	5	0.016718	5	0.004573	Almast by S	111	111	0.10151	0.371147	##########
1213	am	Yerevan	Anoush	##########	3	Armenia	1093485	2990731	Asia	Western As	5	0.016718	5	0.004573	Anoush by	111	111	0.10151	0.371147	##########
1213	am	Yerevan	La traviata	##########	1	Armenia	1093485	2990731	Asia	Western As	5	0.016718	5	0.004573	La traviata	111	111	0.10151	0.371147	##########
1213	ar	Buenos Aire	Carmen	##########	6	Argentina	2891082	44018771	Americas	Latin Ameri	79	0.017947	60	0.020753	Carmen by	526	413	0.142853	0.119494	##########
1213	ar	Buenos Aire	Lucrezia Bo	##########	5	Argentina	2891082	44018771	Americas	Latin Ameri	79	0.017947	60	0.020753	Lucrezia Bo	526	413	0.142853	0.119494	##########
1213	ar	Buenos Aire	I due Figarc	##########	2	Argentina	2891082	44018771	Americas	Latin Ameri	79	0.017947	60	0.020753	I due Figarc	526	413	0.142853	0.119494	##########
1213	ar	Buenos Aire	Cosi fan tut	##########	5	Argentina	2891082	44018771	Americas	Latin Ameri	79	0.017947	60	0.020753	Cosi fan tut	526	413	0.142853	0.119494	##########
1213	ar	Buenos Aire	Le nozze di	##########	6	Argentina	2891082	44018771	Americas	Latin Ameri	79	0.017947	60	0.020753	Le nozze di	526	413	0.142853	0.119494	##########
1213	ar	Buenos Aire	Aleko	##########	4	Argentina	2891082	44018771	Americas	Latin Ameri	79	0.017947	60	0.020753	Aleko by Ra	526	413	0.142853	0.119494	##########
1213	ar	Buenos Aire	Francesca c	##########	4	Argentina	2891082	44018771	Americas	Latin Ameri	79	0.017947	60	0.020753	Francesca c	526	413	0.142853	0.119494	##########
1213	ar	Buenos Aire	La cenerent	##########	5	Argentina	2891082	44018771	Americas	Latin Ameri	79	0.017947	60	0.020753	La cenerent	526	413	0.142853	0.119494	##########
1213	ar	Buenos Aire	Die Frau oh	##########	4	Argentina	2891082	44018771	Americas	Latin Ameri	79	0.017947	60	0.020753	Die Frau oh	526	413	0.142853	0.119494	##########
1213	ar	Buenos Aire	Eugene One	##########	5	Argentina	2891082	44018771	Americas	Latin Ameri	79	0.017947	60	0.020753	Eugene One	526	413	0.142853	0.119494	##########

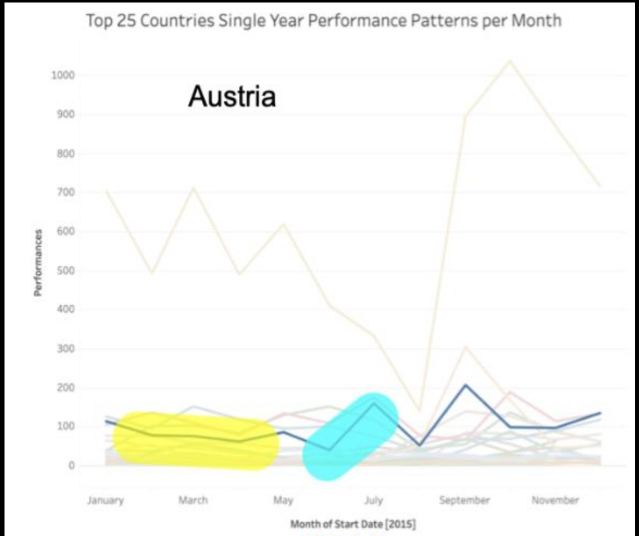
Data Wrangling:

- Merged Population Data per City and Per Country
- Added Country Name, Continent and Sub-regions to data
- Had to manually update cities with multiple spellings from original dataset that did not match with merged dataset
- US cities needed to be manually cleaned due to multiple states having cities with the same names.

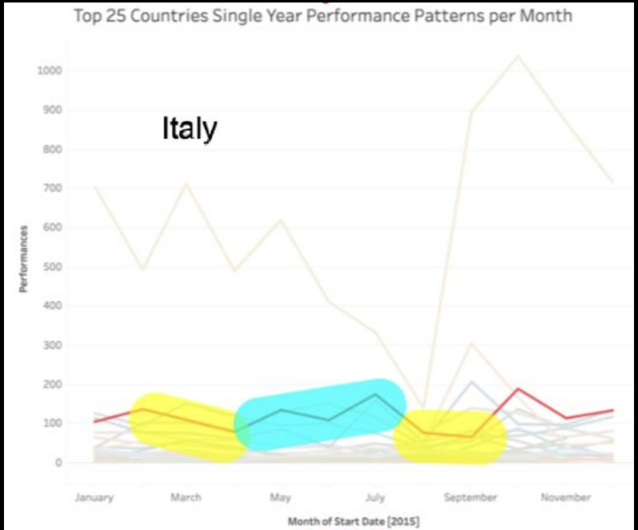
Exploratory Data Analysis (EDA)

These graphs highlight the unique seasonality in Austria and Italy compared to the rest of Europe. This inverse seasonality presents an opportunity for targeted market investment during peak seasons in other countries, while also providing time to explore emerging markets during traditional off-peak seasons, avoiding conflicts with existing opera companies."

- unusual growth from typical patterns
- unusual drop from typical patterns

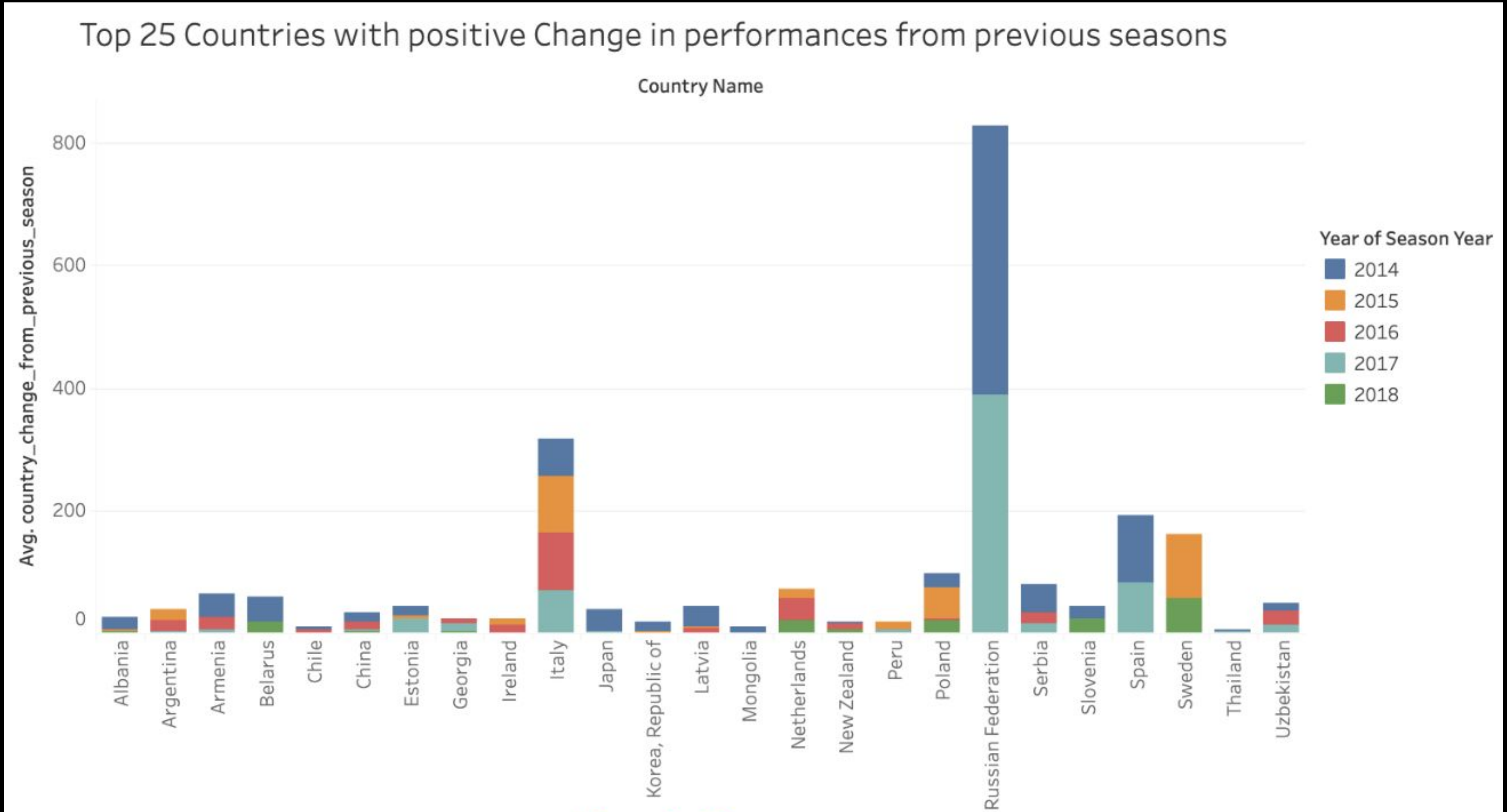


Graph B



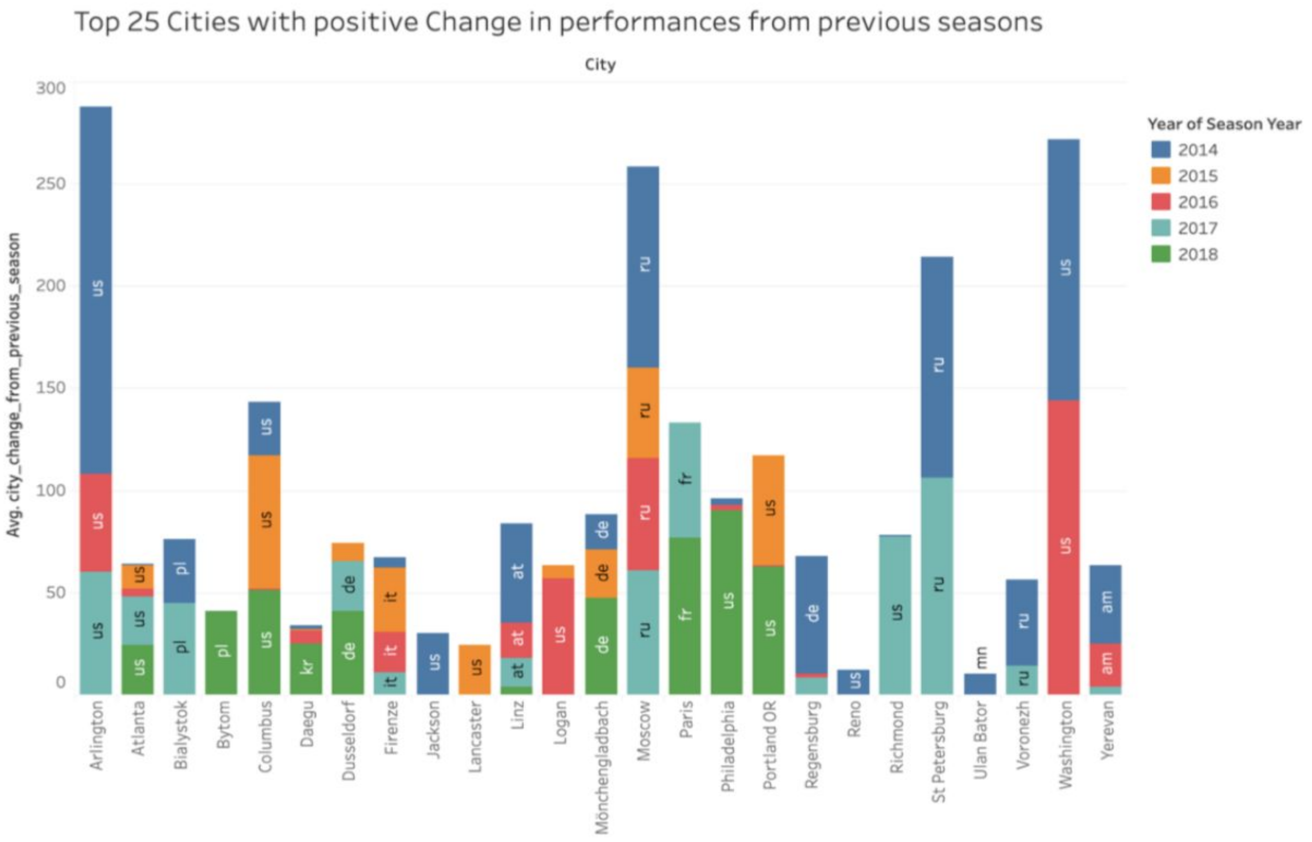
Graph C

EDA continued



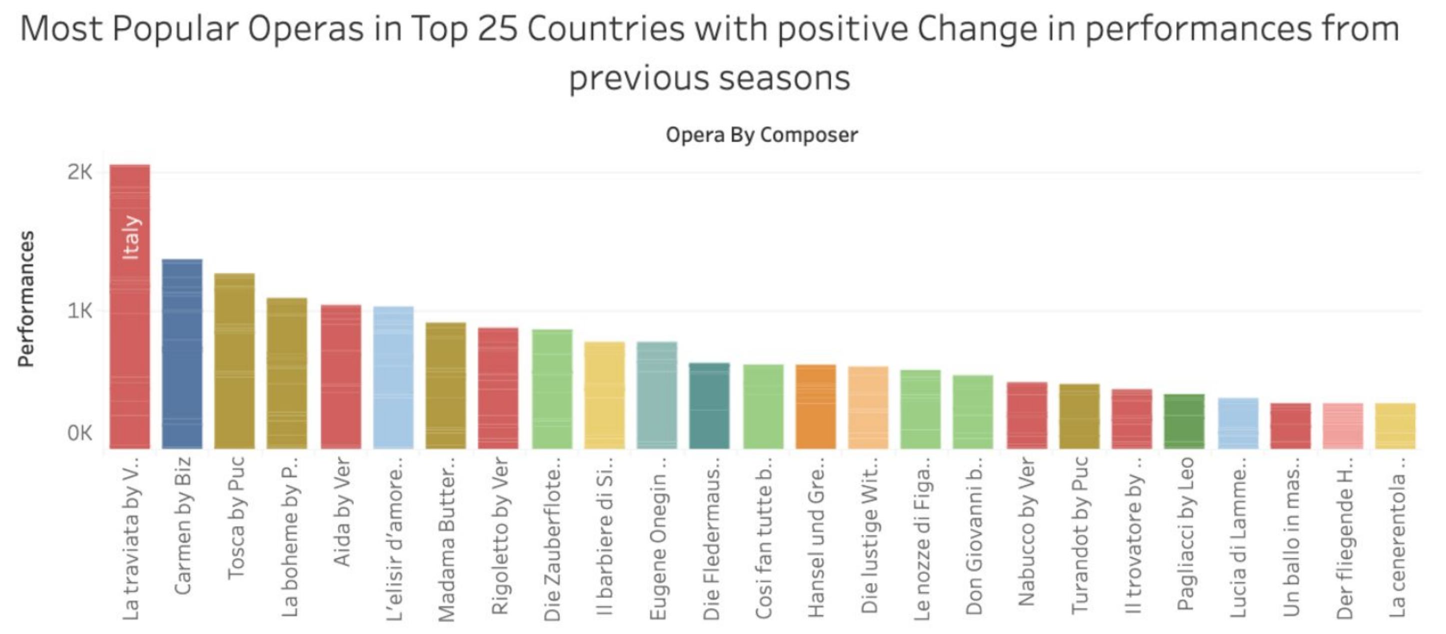
Graph D

EDA continued



Graph E

EDA continued



Graph H

The same operas are in the Top 25 for performed operas with Top 25 Cities (correlating with Graph E), albeit in a different order.

Pre-Processing and Training Data

- Deleted rows not being used for prediction ie. composer, work, composer's data of birth, composers date of death etc.

- Transformed start date into separate datetime objects ie. Month, Day of Week, Year etc.

- Applied one hot encoding for categorical features like Month and Day of Week

- Used linear model to fill in empty values for the change in performances from previous seasons for countries and cities

Modeling - Logistic Regression

How Logistic Regression works

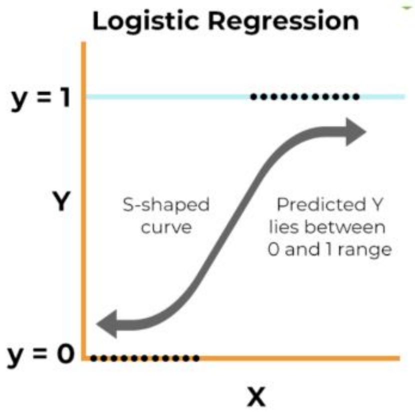


Image Source: https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-logistic-regression/

The Performance Metrics

Country_growth_performance:

- Accuracy: 88.24%
- Precision: 83.10%
- Recall: 84.24%
- f1: 86.33%

City_growth_performance:

- Accuracy: 62.48%
- Precision: 60.75%
- Recall: 62.48%
- f1: 44.95%

The Predicted Countries with Growth

Country Name	predicted_country_growth
Germany	3
France	3
Greece	3
Italy	3
Ireland	3
Hungary	3
United States	3
Croatia	3
Spain	3
Estonia	3

The Predicted Cities with Growth

iso	city	predicted_city_growth
us	Houston	Very High
de	Wiesbaden	Very High
ru	St Petersburg	Very High
hu	Pécs	Very High
ca	Toronto	Very High
hu	Szeged	Very High
de	Regensburg	Very High
dk	København	Very High
ee	Tallinn	Very High
ca	Vancouver	Very High

Modeling - Random Forest Classifier

How Random Forest Classifier works

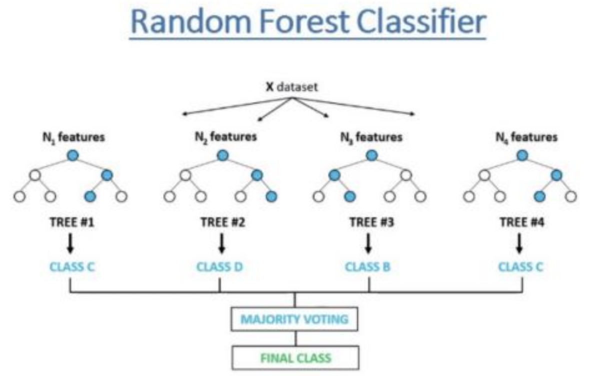


Image Source: https://www.freecodecamp.org/news/how-to-use-the-tree-based-algorithm-for-machine-learning/

The Performance Metrics

Country_growth_performance:

- Accuracy: 97.22%
- Precision: 97.90%
- Recall: 97.22%
- f1: 99.97%

City_growth_performance:

- Accuracy: 93.87%
- Precision: 93.75%
- Recall: 93.87%
- f1: 88.60%

The Predicted Countries with Growth

Country Name	predicted_country_growth
Germany	Very High
Russian Federation	Very High
Slovenia	Very High
Poland	Very High
Serbia	Very High
Sweden	Very High
United States	Very High
Ukraine	Very High
United Kingdom	Very High
Mexico	Very High

The Predicted Cities with Growth

iso	city	predicted_city_growth
us	Lancaster	Very High
us	Denver	Very High
us	Dayton	Very High
de	Landshut	Very High
de	Lubeck	Very High
de	Halle	Very High
us	Houston	Very High
de	Hannover	Very High
de	Heidelberg	Very High
de	Heidenheim	Very High

Modeling - XGBoost Classifier

How Random Forest Classifier works

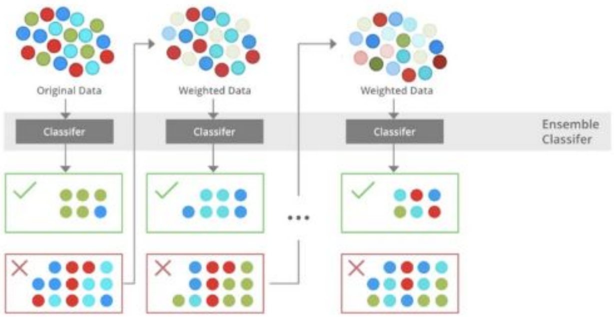


Image Source: https://www.geeksforgeeks.org/xgboost/

The Performance Metrics

Country_growth_performance:

- Accuracy: 90.89%
- Precision: 88.00%
- Recall: 90.89%
- f1: 99.90%

City_growth_performance:

- Accuracy: 92.90%
- Precision: 92.13%
- Recall: 92.90%
- f1: 97.23%

The Predicted Countries with Growth

Country Name	predicted_country_growth
Germany	3
Italy	3
Japan	3
Korea, Republic of	3
United States	3
United Kingdom	3
Poland	3
Latvia	3
Mexico	3
Hungary	3

The Predicted Cities with Growth

iso	city	predicted_city_growth
us	Lancaster	3
de	Greifswald	3
ru	Moscow	3
de	Gera	3
de	Goerlitz	3
se	Stockholm	3
de	Flensburg	3
rs	Beograd	3
ru	Krasnoyarsk	3
ru	Ekaterinburg	3

Conclusions:

In Top Countries (Graph D)

In both Top Countries and Top cities (Graph D and Graph E)

Has Top Cities in Graph E

Country Name	predicted_country_growth	Model
Argentina	3.0	XGB - Country
Australia	3.0	Both
Belgium	3.0	XGB - Country
Brazil	3.0	LR - Country
China	3.0	Both
Croatia	3.0	LR - Country
Czech Republic	3.0	Both
Denmark	3.0	Both
Estonia	3.0	Both
Finland	3.0	LR - Country
France	3.0	Both
Germany	3.0	Both
Greece	3.0	LR - Country
Hong Kong, China	3.0	LR - Country
Hungary	3.0	Both
Ireland	3.0	Both
Italy	3.0	Both
Japan	3.0	Both
Korea, Republic of	3.0	Both
Latvia	3.0	Both
Lithuania	3.0	LR - Country
Mexico	3.0	Both
Netherlands	3.0	Both
New Zealand	3.0	XGB - Country
Norway	3.0	Both
Peru	3.0	XGB - Country
Poland	3.0	Both
Portugal	3.0	LR - Country
Romania	3.0	LR - Country
Russian Federation	3.0	Both
Serbia	3.0	Both
Slovakia	3.0	XGB - Country
Slovenia	3.0	XGB - Country
Spain	3.0	Both
Sweden	3.0	Both
Ukraine	3.0	XGB - Country
United Kingdom	3.0	XGB - Country
United States	3.0	Both

This chart reaffirms the recommendations from the beginning of the presentation.

Questions?