# Applied Data Science Capstone Course – Final Project

## 1. Introduction (Business Case)

Most of the tourists who visit Peru first land in Lima, the capital city, but most of them immediately take an airplane to Cusco, where the famous Machu Picchu is located. This is due to the abundant advertisement that highlights Machu Picchu as the most important touristic spot, reducing the chances of tourists visiting other interesting spots in Lima or in Cusco region too. However, tourists might not have the same interests. For example, some of them might be more interested in visiting historic places such Machu Picchu or similar Inca fortress, while others might find dynamic cities as Lima more exciting to visit. If we compare both cities, based on the website Versus.com [1] that gives unbiased comparisons using objective information from many categories, Lima has a higher average temperature (19.4°C vs. 12°C), has seaside beach, 4 more big sports facilities (stadiums, arenas, etc.), and has a bike-sharing system; on the other side, Cusco has lakes and smaller humidity rate (55.4% vs. 83%).
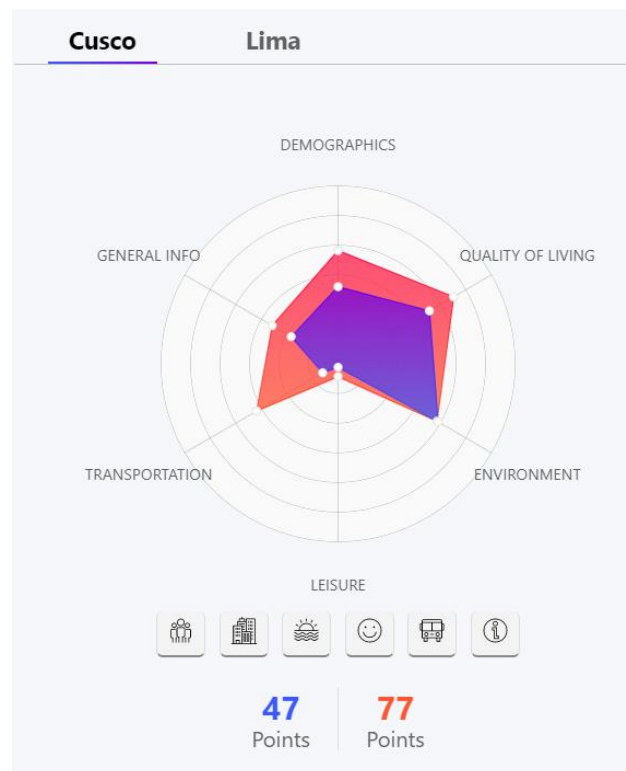


Fig. 1 – Screenshot of comparison between Lima and Cusco from Versus.com
(Cusco in purple, Lima in red)

As we can see, Lima can offer some advantages or may have other attractions than the historical spots located in Cusco. Still, visitors might not be aware of this, trusting blindly in touristic agencies, then getting disappointed when they get to know from other sources that they missed attractions in Cusco or Lima. In consequence, it would be important to explore both cities to find other interesting and beautiful hidden spots around already famous touristic spots. Additionally, it would be important to segment the touristic sites in both cities to create a profile for each cluster, so it would be easier to provide accurate information

to future visitors, helping them to choose where to visit according to their interests. This project aims to fulfill both activities (explore and segmenting), which can be helpful for touristic agencies to extend their packages or programs adding other interesting spots that are near already included sites and create packages according to clients' profiles or interests, resulting in an increased profit. On the other hand, it might result in a higher number of tourists visiting both cities which will help the local government to increase income from touristic activities.

## 2. Data Sources and Usage

Before introducing the list of data sources, it is important to indicate that Peru is divided in regions and each region is divided in provinces, which are also divided in districts. In this scenario, the first data source would be a list of all districts in Peru grouped by province and region [2], which has the format shown in Fig. 2. The website will be scrapped and then using PositionStack API [3] the coordinates of each district will be found.

| Ubigeo ⬍ | Departamento ⬍ | Provincia ⬍ | Distrito ⬍ | Población 2017 ⬍ |
|---|---|---|---|---|
| 010101 | Amazonas | Chachapoyas | Chachapoyas | 32 589 |
| 010102 | Amazonas | Chachapoyas | Asunción | 262 |
| 010103 | Amazonas | Chachapoyas | Balsas | 1136 |
| 010104 | Amazonas | Chachapoyas | Cheto | 642 |
| 010105 | Amazonas | Chachapoyas | Chiliquín | 585 |
| 010106 | Amazonas | Chachapoyas | Chuquibamba | 1781 |
| 010107 | Amazonas | Chachapoyas | Granada | 480 |
| 010108 | Amazonas | Chachapoyas | Huancas | 1258 |
| 010109 | Amazonas | Chachapoyas | La Jalca | 3978 |
| 010110 | Amazonas | Chachapoyas | Leimebamba | 3620 |
| 010111 | Amazonas | Chachapoyas | Levanto | 794 |
| 010112 | Amazonas | Chachapoyas | Magdalena | 852 |
| 010113 | Amazonas | Chachapoyas | Mariscal Castilla | 1367 |
| 010114 | Amazonas | Chachapoyas | Molinopampa | 2176 |
| 010115 | Amazonas | Chachapoyas | Montevideo | 496 |
| 010116 | Amazonas | Chachapoyas | Olleros | 375 |
| 010117 | Amazonas | Chachapoyas | Quinjalca | 769 |
| 010118 | Amazonas | Chachapoyas | San Francisco de Daguas | 295 |
| 010119 | Amazonas | Chachapoyas | San Isidro de Maino | 580 |
| 010120 | Amazonas | Chachapoyas | Soloco | 1224 |
| 010121 | Amazonas | Chachapoyas | Sonche | 247 |
| 010201 | Amazonas | Bagua | Bagua | 27 102 |
| 010202 | Amazonas | Bagua | Aramango | 9765 |
| 010203 | Amazonas | Bagua | Copallín | 4595 |
| 010204 | Amazonas | Bagua | El Parco | 1216 |
| 010205 | Amazonas | Bagua | Imaza | 25 162 |
| 010206 | Amazonas | Bagua | La Peca | 6260 |
| 010301 | Amazonas | Bongará | Jumbilla | 1337 |
| 010302 | Amazonas | Bongará | Chisquilla | 306 |

Fig.  2 - List of all districts in Peru from Wikipedia website (only available in Spanish)

For the first objective, using the Foursquare API we will explore attractions (and their categories) located in the same district where are the popular touristic spots in each city listed by the Ministry of Foreign Trade and Tourism in its report of tourism development elaborated in 2018 [4] [5]. This will show other interesting sites that might be included in touristic packages and be visited while going to the already included famous spots.

For the second objective, the complete list of attractions in each city will be segmented and clustered, resulting in a division for each city composed by clusters with defined characteristics which can become new touristic packages that will be offered targeting specific visitors according to their interests.

**References:**

[1] https://versus.com/en/cusco-vs-lima

[2] https://es.wikipedia.org/wiki/Anexo:Distritos_del_Per%C3%BA

[3] https://positionstack.com/

[4] https://www.mincetur.gob.pe/wp-content/uploads/documentos/turismo/estadisticas/ReporteTurismoRegional/RTR_Cusco.pdf

[5] https://www.mincetur.gob.pe/wp-content/uploads/documentos/turismo/estadisticas/ReporteTurismoRegional/RTR_Lima_2018.pdf