**Project 1**

1. **Project timeline 1**
2. [**Project Proposal**](#_hnxypnkdg8ht) **1**
3. [**Data Cleanup & Analysis**](#_6a53do2dvjw4) **2**
4. **Tezca Solutions Presentation 2**

# **Project timeline**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Session Name** | **Session #** | **Date** | **Day** | **ETA - Status** |
| Project Work | 7.1 | 27/03/2019 | W |  |
| Project Work | 7.2 | 30/03/2019 | S | Datasets ubicados y normalizados |
| Project Work | 7.3 | 01/04/2019 | M | Regresiones, modelos estadísticos y focos |
| Project Work | 8.1 | 03/04/2019 | W | Pizza y chelas |
| Presentation | 8.2 | 06/04/2019 | S |  |
| Project Work | 8.3 | 08/04/2019 | M |  |

# **Project Proposal: Siguientes puntos de crecimiento a nivel geolocalización.**

# Puntos clave de MX - ciudad/estado

# Niveles de correlación entre datasets

* + Rangos de edades
  + Categorías
  + Triángulo de crecimiento de ecommerce
    - México - Monterrey - Guadalajara

**Data:**

* + Resultado: rate de crecimiento
  + Densidad poblacional
  + Industrialización en México
  + Números de supermercados/construcciones/cafés/restaurantes
  + Escolaridad de población: inscritos en universidad
  + Digitalización: penetración de mercados digitales
  + Facturación (mayor volumen de transacciones)
  + Tráfico de internet
* redes cura - red de fibra optica

**Data Sources**

* <https://www.kaggle.com/skillsmuggler/amazon-ratings>
* alexa.com, playbooks
* bloomberg
* <https://www.gapminder.org/data/>

|  |
| --- |
| **Suggested Data Sources** |
| * [data.world](https://data.world/) * [Kaggle](https://www.kaggle.com/) * [Data.gov](https://www.data.gov/) * [Public APIs](https://github.com/abhishekbanthia/Public-APIs) * [Awesome-APIs List](https://github.com/Kikobeats/awesome-api) * [Medium APIs List](https://medium.com/@benjamin_libor/a-curated-collection-of-over-150-apis-to-build-great-products-fdcfa0f361bc) |

# **2. Data Cleanup & Analysis**

|  |  |  |
| --- | --- | --- |
| **Tool** | **Task** | **ETAs** |
| 1. Pandas 2. Jupyter Notebook 3. Matplotlib 4. PNG images 5. Use 1 API (option) 6. Write up | 1. clean and format datasets 2. describing the data exploration / cleanup / final data analysis 3. 6-8 visualizations of your data ( 2 per "question" asked) 4. For ppt 5. Optional Api 6. Summarizing + heading for each "question" & short description of what you found and any relevant plots. |  |

# **3. Presentation**

|  |  |
| --- | --- |
| **Presentation Requirements** | **Presentation Suggestion** |
| * 8-10 min. long * Describe core message/hypothesis. * Describe questions/motivations * Summarize where and how you found the data you used * Describe data exploration and cleanup process (+ Jupyter Nb) * Describe analysis process (+ Júpiter Nb) * Summarize conclusions   + numerical summary (what data did your analysis yield)   + visualizations (plots of the final analysis data) * Implications of your findings. | * Title Slide: project name & members * Motivation & Summary Slide   + Core message/hypothesis   + Questions and reasoning   + Good answers? Summary * Questions & Data   + Elaborate on the questions you asked & data needed * Data Cleanup & Exploration   + Describe process   + Insight/problems/resolutions encountered * Present and discussion * Data Analysis steps   + Answer all question of the proposal   + Present interesting figures * Discussion: Expectations, conclusions? * Post Mortem: difficulties/solutions, what would u do differently * Q&A |

**Data interpretation**

* Datasets with time intervals (population growing, enrolled in school, population under 30 years) can define a trend.
* Current situation can provide a reference (internet penetration, bank data, financial products, ).

Data sets

**Can provide a trend**

Population growing

Enrolled in school

Population under 30 years

New starbucks (starbucks index, gentrification indicator)

Human Development Index

**Actual status**

Bank data

Financial products

Internet penetration

Mobil plans

Web traffic (searching)