



UNLP
Facultad de Informática
Tecnologías Aplicadas para Business Intelligence
Curso 2020

Tecnologías Aplicadas para Business Intelligence

Trabajo Final 2020

Durante el desarrollo de la materia fuimos viendo los fundamentos del modelado dimensional, las ventajas que tiene como fuente de análisis de información y generación de informes y visualizaciones.

Pusimos en práctica estos conocimientos con algunas herramientas muy útiles y vimos además de forma acotada otros temas y herramientas, que si bien se abordan brevemente, les dan pautas para avanzar en su estudio y así extender sus habilidades.

El trabajo final de la materia, es de libre elección, siempre que contenga como parte de su desarrollo parte de lo expuesto en la materia.

Siempre recomendamos que puedan tomar algo que les interese, ya sea porque es parte de su trabajo, o porque les sea de alguna manera útil.

Deberán inicialmente presentar una propuesta de trabajo final de no más de una carilla donde informen lo siguiente:

- Objetivo del proyecto
- Alcance de mínimo y posibles extensiones

Recuerden que disponen de 1 mes aproximadamente para realizar el desarrollo y una breve exposición para comunicarlo al resto de sus compañeros.

Durante ese tiempo pueden realizar consultas, ya sea por el foro o en encuentros virtuales

Ideas principales

Como concepto principal deben recopilar información de distintas fuentes e integrar en una base de datos dimensional para realizar informes y visualizaciones que les permitan aportar valor para la compresión del estado de las cosas y eventualmente ser de utilidad para toma de decisiones.

Datos

Pueden tomar datos de portales de datos abiertos (algunos de ellos se dejan de referencia), o utilizar alguna herramienta de scrapping para obtenerlos de algún sitio público, también pueden utilizar datos propios, siempre y cuando no violen principios de privacidad de los mismos.

Deberán explicitar de qué forma van a integrar los datos y diseñar un modelo dimensional que sea útil a sus propósitos.

Comunicación

Basado en los datos, deberán crear algún informe, visualización o desarrollo que sea de utilidad para la toma de decisiones.

Alternativas

Si tienen en mente realizar algo distinto, y su proyecto contempla los contenidos básicos de la materia no duden en proponerlo.

Algunas ideas

Las siguientes son solo algunas sugerencias por si no se les ocurre qué hacer.

- **COVID19**
- **Apertura y Cierres de empresas en Argentina**
- **Empleo**
- **Propiedades, valores y características de las ofertas**
- **Tendencias de Consumo**
- **Cualquier otro a elección**

Calendario de actividades

20/11: Presentación de la propuesta inicial a través de Moodle.

Entrega de documento especificando el objetivo, e idea general.

Alcance mínimo y posibles extensiones.

15/12: Presentación. Clase Obligatoria.

Para presentar en clase el trabajo realizado por cada grupo.

Exposición de 20 a 25 minutos

Pueden incluir un video demo online

Subir presentación a Moodle

Referencias

Portales de datos abiertos

- <https://datos.gob.ar/>
- <http://datos.salud.gob.ar/dataset?groups=covid-19>
- <https://www.buenosaires.gob.ar/coronavirus/datos>
- <https://datos.laplata.gov.ar/>
- <https://datos.bancomundial.org/>
- <https://www.gapminder.org/tag/download-data/>
- <https://www.properati.com.ar/data/>
- <https://www.kaggle.com/>
- <https://github.com/caesar0301/awesome-public-datasets>
- <https://github.com/datosgobar>
- Cualquier otro que encuentren

Herramientas para la integración de los datos

- Pentaho Data Integration
- Python + Pandas <https://pandas.pydata.org/>
- Cualquier otra herramienta o lenguaje que les sea de utilidad
- Apache Airflow <https://airflow.apache.org/>
- Prefect <https://docs.prefect.io/>

- Luigi <https://github.com/spotify/luigi/tree/master/examples>
- Apache Nifi <https://nifi.apache.org/>
- Apache Spark <https://spark.apache.org/>
 - pyspark
 - databricks <https://databricks.com/youtube-lista>
- Apache Kafka <https://kafka.apache.org/>
- Apache Druid <https://druid.apache.org/>
- Más info acá:
 - <https://github.com/igorbarinov/awesome-data-engineering>
 - <https://awesomedataengineering.com/>
- Scrapping
 - scrapy
 - BeautifulSoup
 - puppeteer
 - rvest

Herramientas para generar visualizaciones

- Grafana provisto por la cátedra
- D3.js <https://d3js.org/>
- Highcharts <https://www.highcharts.com/>
- OpenStreetMaps <https://www.openstreetmap.org/about>
- CartoDB <https://carto.com/>
- LeafletJS <https://leafletjs.com/>
 - folium (python wrapper) <https://github.com/python-visualization/folium>
 - react-leaflet: <https://react-leaflet.js.org/>
- Python
 - Matplotlib <https://matplotlib.org/>
 - Seaborn <https://seaborn.pydata.org/>
 - Plotly <https://plotly.com/>
 - Streamlit <https://www.streamlit.io/>
 - Bokeh <https://bokeh.org/>
 - geopandas <https://geopandas.org/>
 - deck.gl y pydeck <https://deck.gl/docs>
- Google Data Studio <https://datastudio.google.com/>

- Tableau <https://www.tableau.com/>
- PowerBI <https://powerbi.microsoft.com/>
- Qlik <https://www.qlik.com/es-es/>
- Google Maps <https://developers.google.com/maps/?hl=es-419>
- Shiny R Studio. <https://shiny.rstudio.com/tutorial/>
- R <https://rstudio.com/>