USC Viterbi
School of Engineering
Information Technology Program

ITP 487 – Enterprise Data Analytics

Units: 4
Fall 2022, Noon–1:50PM MW

Location: ZHS 159

Instructor: Mike Lee
Contact Info: mikelee@usc.edu
Office Hours: bit.ly/professorlee

Teaching Assistants:

- Brett Evenhouse (bevenhou@usc.edu) – Lead
- Indu Gaddamadugu (gaddamad@usc.edu)
- Aryan Ahuja (aryanahu@usc.edu)
- Amy Jiang (amyj@usc.edu)

See bit.ly/professorlee for latest info

IT Help:
USC IT (ITS): https://itservices.usc.edu/contact/
Viterbi IT: https://viterbi.usc.edu/resources/vit/contact-us.htm

Course Description

While the increased capacity and availability of data gathering and storage systems have allowed enterprises to store more information than ever before, most organizations still lack the ability to effectively consolidate, arrange, and analyze this vast amount of data. Digital transformation using data analytics techniques has become a highly sought-after skill in business, engineering, services, science, health, and other industries.

This course will explore the theory and practice of the following areas:

- Enterprise Structure and Decision Making
- Enterprise Data Warehouses
- Data Analytics used by Enterprises
- USC Applied Data Analytics Methodology (ADAM)

Learning Objectives

After completing the course, students will be able to

- Understand the organizational structure of enterprises (large organizations)
- Understand how enterprises make major technology decisions
- Define enterprise data analytics and its drivers
- Describe the components of an enterprise data warehouse
- Model the relational database required for an enterprise data warehouse
- Extract, cleanse, consolidated, and transform heterogeneous data into a single enterprise data warehouse
- Explore any data set and apply a repeatable approach to data analytics to gain relevant insights
- Apply data analytics techniques that is in demand by enterprises

Prerequisite(s): ITP 320 or ITP 249

Remote Attendance

This course does not support remote attendance. Lectures will not be recorded or available on Zoom, there are short in-person individual/group activities during many class meetings and exams are in-person.

Course Notes

All course materials will be made available through Blackboard. These include:

- Lecture slides
- In-class exercises
- Homework assignments
- Readings
- Software details and instructions for accessing Viterbi Virtual Lab
- Grades and feedback
- Office hours
- Online discussion forums will be used for out-of-class discussions

Announcements made in class and content posted in Blackboard will supersede the contents of this syllabus.

USC Technology Support Links

Zoom information for students
Blackboard help for students
Software available to USC Campus

Technological Proficiency and Hardware/Software Required

The assignments for this class will include both reading assignments as well as hands-on computer assignments. Students must bring their laptop computers (phones/tablets are not sufficient) to lecture sessions to participate in hands-on activities. Students will be given tutorials to gain familiarity with software tools.

Most of the enterprise software required for the class is Windows based or delivered via the cloud. The software will be provisioned through the Viterbi Virtual Lab, Amazon Web Services, Google Cloud, and/or installed your computer. Specifically, students will be using:

- SAP BW/4HANA (cloud)
- Eclipse for SAP BW Modeling (installed locally or Viterbi Virtual Lab)
- SAP Analysis for Microsoft Excel (installed locally or Viterbi Virtual Lab)
- Amazon Web Services/RDS (cloud)
- Google Big Query (cloud)
- Relational Database/SQL (cloud & Viterbi Virtual Lab)
- Google Colab/Python/Pandas (cloud)
- Github (cloud)
- Microsoft Power BI* (Viterbi Virtual Lab)
- Python/Pandas
- SQL

* bonus exercises based on in class data sets

VITERBI VIRTUAL LAB – VMWARE VDI

Some software can also be accessed via Virtual Desktop by logging into the General Desktop at: http://mydesktop.vlabs.usc.edu. If prompted enter http://mydesktop.vlabs.usc.edu as the VDI server. See blackboard for additonal instructions on installing.

Alternatively, you can install the required software on your Windows machine (no support will be provided). Instructions will be posted on Blackboard.

Readings and Supplementary Materials

Reading and supplementary materials will be announced in class and published on Blackboard.

Optional: Practical Analytics, Nitin Kale and Nancy Jones, Second Edition, Epistemy Press 2020
http://store.epistemypress.com/books/analytics.html

Description and Assessment of Assignments

Homework: Most homework is computer based. Homework should be turned in to Blackboard. Grading will be based on completeness, accuracy, and timeliness. Feedback will be provided through Blackboard. These are individual effort assignments. ***One homework assignment will be dropped (lowest score) from your grade calculation.***

In-Class Exercises: are guided Q&A and hands-on exercises that are used to spark additional discussion and deeper understanding of the materials and concepts before the student leaves the class. Announcement of in-class exercises may or may not be given prior to the class. In-class exercises can be team or individual exercises. The score used for grading is the percentage of in-class exercises completed and turned in in-class. ***Two in-class exercises will be dropped (lowest scores) from your grade calculation.***

Exams: Each exam will be comprised of 1) in-person and in-class multiple-choice "scantron" sheets during class-time and 2) take-home project that you will have several days to complete. Details will be posted on Blackboard.

Final Project: Final project is an individual summative assignment where you will be applying skills that you have learned through the semester.

Grading Breakdown

Homework	30%
In-Class Exercises	5%
Exam I	20%
Exam II	20%
Final Exam	20%
Final Project	5%
TOTAL	100%

Grading Scale

Final grades represent how you perform in the class relative to other students. Historically, the average grade for this class is about a 3.4.

Grading Timeline

Grading will typically be completed 7 days after submission. Any variations will be announced in class or on blackboard.

Policies

Students are expected to attend and participate in lecture discussions, in-class exercises, and team meetings.

Assignments turned in late will have 25% of the total points deducted from the graded score for each late day.

No make-up exams (except for documented medical or family emergencies) will be offered. If they will not be able to attend an exam due to an athletic game or other valid reason, then they must coordinate with the instructor before the exam is given. They may arrange to take the exam before they leave, with an approved university personnel during the time they are gone, or within the week the exam is given. If students do not take an exam, then they will receive a 0 for the exam. Accommodations religious observance must be arranged with the Professor at least two weeks before the exam.

If students need accommodations authorized by OSAS (Office of Student Accessibility Services), notify the instructor at least two weeks before the exam. This will allow time for arrangements to be made.

Sharing of course materials outside of the learning environment

SCampus Section 11.12(B)

Distribution or use of notes or recordings based on university classes or lectures without the express permission of the instructor for purposes other than individual or group study is a violation of the USC Student Conduct Code. This includes, but is not limited to, providing materials for distribution by services publishing class notes. This restriction on unauthorized use also applies to all information, which had been distributed to students or in any way had been displayed for use in relationship to the class, whether obtained in class, via email, on the Internet or via any other media. (See Section C.1 Class Notes Policy).

Course Schedule

Week	Date	Lecture Topic(s)	See Blackboard for Due Dates
1	8/22	Course Introduction • What is an Enterprise? • What is Enterprise Data Analytics? • USC Applied Data Analytics Methodology (ADAM) • Use Cases • Career Relevance • Course Overview	
	8/24	Enterprise Structure • Structure of enterprises • CIO reporting structure • Technology decision making • Components of a strategy • Strategy tools – Gartner Magic Quadrant Enterprise Technologies and Cloud Services • What is Enterprise Cloud Services? • Cloud Service Types: IaaS, PaaS, SaaS, and BPaaS • Need for the Cloud • Enterprise Considerations (security/legal) • Enterprise Cloud Landscape • Cloud Services Used in This Class	
2	8/29	Data Concepts • Structured vs Unstructured Data • Type vs Instance • Physical vs Virtual • Data Storage Component • Master Data vs Transactional Data • Big Data vs Enterprise Data Relational Database for Analytics • Relational Database for Analytics vs non-Analytics • Data sets: Narratives, Data dictionary & ER diagrams • Entity relationship diagram (ERD or ER Diagram) • Crows foot notation • Entity, relation, key, attributes, relationships, cardinality, referential integrity constraint, normalization Data Modeling – Lucid Chart • Online ER Diagram Creation • Entities, Key, Field, Type • Relationships & Crows Foot Notation • Exporting ERD to DDL *In-Class: ER Diagram (LucidChart) – design ER diagram in LucidChart and export to SQL DDL.*	HW1: ER Diagram – you will create a data dictionary and ER diagram from narratives TAKE HOME: Install MySQL Workbench on Laptop

	8/31	SQL for Analytics (part 1) • SQL for Analytics vs non-Analytics • Data Manipulation Commands • CREATE, DROP, INSERT, UPDATE, & DELETE • Creating and Modifying Schemas: CREATE SCHEMA • Loading Data: USE & INSERT • Querying: SELECT Amazon Web Services (AWS) / RDS Overview • Amazon Web Services & Relational Data Service • Creating a Database • Creating and Modifying Schemas • Loading Data • Querying *In-Class: Build Your Lab (AWS/RDS) – you will be creating a database in AWS/RDS, building a simple schema, and testing connectivity from MySQL workbench to your database*	
3	9/5	**NO CLASS – LABOR DAY**	
	9/7	SQL for Analytics (part 2) • Data Set: Narrative, Data dictionary & ER diagram • SELECT & JOIN • Loading Data • Querying *In-Class: Create & Load Schema in AWS/RDS – using data set from lecture*	HW#2 - AWS/RDS – you will be implementing the ER diagram that you previously designed in LucidChart in the lab the you previously built int AWS/RDS. You will also load and query the data using SQL.
4	9/12	Google Cloud Platform, Big Query, and Shopify • Overview of Google Cloud/Big Query • Differences between AWS & Google Cloud for Analytics • Creating and Modifying Schemas • Loading Data • Querying *In-Class: Build Your Lab (Google Big Query) you will be creating simple tables in Google Big Query and performance tests*	
	9/14	Google Cloud Platform, Big Query, and Shopify • Online Sales & Distribution (small business example) • Overview of Shopify • Overview of Data Set (Shopify Data) *In-Class: Design, Load, & Analyze Shopify Data – you will add additional data into Google Big Query in preparation for your HW assignment.*	HW#3 – Google Big Query/SQL (Shopify) – this will be a continuation of the work that you have performed in in-class exercises. You will be using Google Big Query to perform more complex analysis of Shopify data

5	9/19	Data Warehouse Concepts
		• What is a Data Warehouse?
		  ○ Transactional (OLTP) vs Data Warehouse Systems (OLAP)
		  ○ Need for Data Warehouse
		• Importance of Master Data
		• Data Warehouse Components
		• Data Warehouse Process
		• Multi-Dimensional Data
		  ○ Multi-Dimensional Modeling
		  ○ Multi-Dimensional Data Storage
		  ○ Tabular vs multi-dimensional data
		• Star Schema
		  ○ Fact tables (transactional data)
		  ○ Dimension tables (master data)
		• Design Star Schema from Transactional Systems
		Data Warehouse Implementation
		• Identify, Collect, & Clean Data: Understand the Data
		• Model: Design Star Schema & Load Data
		• Analyze: Query Star Schema	

	9/21	Data Warehouse Implementation • Identify, Collect, & Clean: Narrative, Data dictionary & ER diagram • Model: Dimension tables (master data) • Model: Fact tables (transactional data) *In-Class: Design and load Star Schema Datawarehouse (Lucid Chart, AWS RDS, MySQL Workbench)*	HW#4 – Star Schema Warehouse – you will load transactional DDL files, design star schema, load the star schema from tables from the transactional tables using staging tables.

6	9/26	Enterprise Data Warehouse Fundamentals
		• What is an Enterprise Data Warehouse?
		• Additional Enterprise Needs
		  ○ Historization (Time Dependency)
		  ○ Language Dependency
		  ○ Multi-currency/unit of measure
		• Snowflake Schema
		  ○ Fact Tables
		  ○ Dimensional Tables
		  ○ Master Data Tables
		• Master data tables
		  ○ Attributes – Display, Navigational
		  ○ Texts
		  ○ Hierarchies
		• Difference between star schema and snowflake schema	

	9/28	Exam I Review (and take home assigned)	
	10/3	***Exam I (10/3)***	

7	10/5	Enterprise Data Warehouse Implementation
		• Why SAP?
		• SAP HANA Database
		• SAP BW/4HANA
		• SAP BW Key Components
		• Data Warehouse Components vs SAP BW Key Components
		• Other Enterprise Data Warehouses
		
		InfoObjects - SAP's Data Warehouse Catalog
		• Enterprise-Wide Definition
		• Characteristics
		  ○ Master Data
		  ○ Physical Storage
		  ○ Creating Characteristics
		  ○ Time Dependency (Historization)
		  ○ Language Dependency
		  ○ How Used
		• Key Figures
		  ○ Creating Key Figures
		  ○ Multi-Currency
		  ○ Unit of Measure
		  ○ Standard and Exception Aggregation
		  ○ How Used	

8	10/10	Loading Master Data: Characteristics • Moving data through the data warehouse • ETL • Process chains • Master data loading into characteristics InfoProviders: Data Containers/Views • Advanced Data Stores (ADSOs) • Defining an ADSO • Composite providers • Defining a composite provider • Changing output	HW#5 - InfoObjects
	10/12	Loading Transactional Data: Fact Table Loading • Source systems • Data Sources • Extractors for data (APIs etc.) • Mapping of fields • Transformation rules • Data cleansing and harmonization • Transactional data loading into ADSOs	HW#6 – InfoProviders

9	10/17	Data Analyst: Queries
		• Enterprise Analyst Roles & Organization
		• Query basics
		• Query designer
		  ○ Sheet definition
		  ○ Filter
		  ○ Free characteristics
		  ○ Conditions
		  ○ Exceptions
		• Calculated key figures / formulas
		• Navigational and display attributes
		• Currency conversion
		In-Class: Queries
			HW#7 - Queries

	10/19	Business Analyst: Analysis for Excel • Enterprise software nomenclature • Enterprise analyst roles & data organizations • Analysis basics • SAP BW Query vs Analysis for Excel • Dimensions • Members • Hierarchy • Measures • Filter by member • Filter by measure • Conditional formatting *In-Class: Analysis for Excel*	HW#8- Analysis for Excel
10	10/24	USC Applied Data Analytics Methodology (ADAM) • Importance of an Approach • Methodology: Identify, Collect, Clean, Model, Analyze, Publish • Toolkits • Use Cases	
	10/26	Exam II Review	
11	10/31	***Exam II (10/26)***	Bonus HW: Microsoft Power BI – you will be using Power BI to connect to SAP BW to create a dashboard.
	11/2	USC ADAM + Python Toolkit • Google Colab Overview • Notebook vs Runtime • Connecting to a Runtime • Temporary Modules / Files • Github • Python & Libraries • Other sources of data *In-Class: Build Your Lab (Google Colab & Github) - you will be loading data from Github into Google Colab to perform simple analysis. Additional capabilities include loading data from AWS/RDS, Google Big Query, and SAP BW.*	

12	11/7	Pandas for Analytics (part 1)
		• What is Pandas for Analytics?
		• Data Set Overview
		• Data Storage Components
		  ○ Data Frames: Tabular Data
		  ○ Series: Column of Data
		• Loading/saving data
		• Viewing data
		  ○ Shape
		  ○ Selection
		• Cleaning data
		  ○ Remove
		  ○ Format
		  ○ Reshape
		  ○ Lamba
		• Model
		  ○ Calculated columns
		  ○ Merge
		  ○ Grouping
		• Analysis
		
		*In-Class Exercise: Data Set: Wordle**
			HW#9 - Analytics

	11/9	Pandas for Analytics (part 2) – continuing experiential learning using a more complex data set
		• Viewing data
		  ○ Shape
		  ○ Selection
		• Cleaning data
		  ○ Remove
		  ○ Format
		  ○ Reshape
		  ○ Lamba
		• Model
		  ○ Calculated columns
		  ○ Merge
		  ○ Grouping
		• Analysis
		
		*In-Class Exercise: Data Set: NBA Stats 2021-22 Regular Season**	

13	11/14	Data Visualization - Charts • Matplotlib vs Alternatives • Histogram • Bar & Stacked Bar • Line • Area & Area Stacked • Pie & Donut • Scatter & Scattered Bubble *In-Class Exercise: Data Set: Formula 1**	HW#10 – Data Visualization

	11/16	ITP 487 Student Survey – Fall 2022 • Data Set Overview: Student Survey • Analysis Techniques *In-Class Exercise: Data Set: Student Survey**	
14	11/21	TBD – Semester Specific Topic	
	11/23	**NO CLASS - THANKSGIVING**	
15	11/28	Final Exam Review Final Project Overview • Data Set Options • Rubric • Approach / Proposal	Assign Final Project
	11/30	Course Wrap Up & Questions • Industry & Discussions • Bonus Projects	
Finals Week		***Final Exam (see USC finals schedule)*** Final Project Due (see Blackboard)	

* Data sets change each semester. Data set listed is the possible data set that will be used.

Statement on Academic Conduct and Support Systems

Academic Conduct:

Plagiarism – presenting someone else's ideas as your own, either verbatim or recast in your own words – is a serious academic offense with serious consequences. Please familiarize yourself with the discussion of plagiarism in SCampus in Part B, Section 11, "Behavior Violating University Standards" policy.usc.edu/scampus-part-b. Other forms of academic dishonesty are equally unacceptable. See additional information in SCampus and university policies on scientific misconduct, https://ooc.usc.edu/research-compliance/scientific-integrity/.

Support Systems:

Counseling and Mental Health - (213) 740-9355 – 24/7 on call
studenthealth.usc.edu/counseling
Free and confidential mental health treatment for students, including short-term psychotherapy, group counseling, stress fitness workshops, and crisis intervention.

National Suicide Prevention Lifeline - 1 (800) 273-8255 – 24/7 on call
suicidepreventionlifeline.org
Free and confidential emotional support to people in suicidal crisis or emotional distress 24 hours a day, 7 days a week.

Relationship and Sexual Violence Prevention Services (RSVP) - (213) 740-9355(WELL), press "0" after hours – 24/7 on call
studenthealth.usc.edu/sexual-assault
Free and confidential therapy services, workshops, and training for situations related to gender-based harm.

Office of Equity and Diversity (OED) - (213) 740-5086 | Title IX – (213) 821-8298
equity.usc.edu, titleix.usc.edu
Information about how to get help or help someone affected by harassment or discrimination, rights of protected classes, reporting options, and additional resources for students, faculty, staff, visitors, and applicants.

Reporting Incidents of Bias or Harassment - (213) 740-5086 or (213) 821-8298
usc-advocate.symplicity.com/care_report
Avenue to report incidents of bias, hate crimes, and microaggressions to the Office of Equity and Diversity |Title IX for appropriate investigation, supportive measures, and response.

The Office of Disability Services and Programs - (213) 740-0776
dsp.usc.edu
Support and accommodations for students with disabilities. Services include assistance in providing readers/notetakers/interpreters, special accommodations for test taking needs,

assistance with architectural barriers, assistive technology, and support for individual needs.

USC Campus Support and Intervention - (213) 821-4710
campussupport.usc.edu
Assists students and families in resolving complex personal, financial, and academic issues adversely affecting their success as a student.

Diversity at USC - (213) 740-2101
diversity.usc.edu
Information on events, programs and training, the Provost's Diversity and Inclusion Council, Diversity Liaisons for each academic school, chronology, participation, and various resources for students.

USC Emergency - UPC: (213) 740-4321, HSC: (323) 442-1000 – 24/7 on call
dps.usc.edu, emergency.usc.edu
Emergency assistance and avenue to report a crime. Latest updates regarding safety, including ways in which instruction will be continued if an officially declared emergency makes travel to campus infeasible.

USC Department of Public Safety - UPC: (213) 740-6000, HSC: (323) 442-120 – 24/7 on call
dps.usc.edu
Non-emergency assistance or information.

Office of the Ombuds - (213) 821-9556 (UPC) / (323-442-0382 (HSC)
ombuds.usc.edu
A safe and confidential place to share your USC-related issues with a University Ombuds who will work with you to explore options or paths to manage your concern.