Data Analytics OL Notes

What is data analytics?

Data analysis techniques that can understand, describe, and predict meaningful trends, patterns, and behaviors from the data deluge often referred to as “Big Data.” Those working in the field of Data Analytics are responsible for the aggregation, storage, and management of very high volumes of data and the subsequent analysis of this data to determine trends, construct descriptive, predictive, and prescriptive models for forecasting and nowcasting, and to aid in strategic decision-making across heterogeneous domains including healthcare, finance, manufacturing, biopharmaceuticals, marketing, education, and in both the private and public sectors.

What are the prereqs to be successful in this program?

DAN prep:

A student would need one undergraduate course in statistics. Since there are programming-based assignments in some of the required courses, and some electives that require coding background, we require that students be familiar with programming concepts. There is no need to be a coder, just one semester of programming. Applicants need more than R, SAS, or MATLAB programming. The course we recommend uses Java, but Python is also fine.

If you lack the programming or stats, you could take a 3-credit undergraduate intro to programming or stats class at a community college (in an OO or imperative language....Java, Python, C++, C# are fine) or Penn State's online class, IST 140 or STAT 200 as a non-degree student.  To register as a non-degree student, you should contact the World Campus Registrar at [registration@worldcampus.psu.edu](mailto:registration@worldcampus.psu.edu). You may also call them directly at 800-252-3592.

BAN prep

A student would need one undergraduate course in statistics. For the business analytics track, no proficiency in coding is required.

MAN prep

The program requires 1 semester of undergrad stats and comfort with quantitative data and IT skills. The marketing analytics concentration requires basic understanding of marketing gained through either a marketing course (undergrad or grad) and/or marketing experience. You need to understand fundamental marketing concepts and strategies - such as brand equity, brand management, product and pricing strategies, customer management (e.e. acquisition, churn, upsell/cross-sell), customer satisfaction.

What are the differences between the base and the big data option?

|  |  |  |
| --- | --- | --- |
|  | Base Program | Big Data Option |
| Aim | Provides a broad foundation in data analytics, covering various aspects such as data mining, predictive analytics, and data management; students learn to collect, classify, analyze, and model data using tools like R, Python, SQL, and Hadoop. | Provides skills to design and management of large-scale databases and real-time data analytics; students gain expertise in big data storage technologies, NoSQL databases, data warehousing, and real-time analytics. |
| Target | Professionals seeking to advance their career primarily in data analytics | Professionals seeking to advance their career primarily in big data management and architecting big data systems. |
| Jobs | Graduates are prepared for roles such as data analyst, data scientist, and analytics manager. | Graduates are prepared for roles such as big data engineer, database administrator, and systems architect. |

Can I use a Mac?

We do not recommend students use MacOS, no, although many do still try. The issue is that our faculty cannot provide technical support when incompatibilities arise with the prescribed toolsets. They general provide as much assistance as they can, but they don’t have multiple machines at their disposal. This is a particular problem given the changes in chipsets in Macs – from Intel-based chips, to the M1, and now M2 chip machines.

What is the average class size?

The average class size is 24.

What statistics software is taught in the MAN option?

In required courses, Minitab and R. In electives SPSS and SAS.

Can I decide if I want the business analytics classes later?

The difference is 3 courses and students can apply to change from one to another, but the decisions are made by different admissions committees, so entrance into one does not guarantee entrance into another. Only 3 credits of the nine required for an option can be applied to another option, so while students can move between options, it may mean that some courses don’t apply to the new option. The business analytics option courses must be taken in order (although not necessarily in consecutive semesters) whereas the base option courses are all scheduled every semester.

How many start dates per year?

We offer three: fall, spring, and summer.

Who are the faculty?

Here are links to some of the faculty involved in the development and delivery of the Data Analytics program. They include authors of the INFORMS CAP certification exam, nationally and internationally renowned experts in analytics, data mining, and information retrieval and storage. There are a number of “Big Data” initiatives and labs at Penn State and many of our faculty are deeply involved in one or more of them. One example is the Big Data Lab at Penn State Great Valley which is, among other things, mining social media posts regarding institutions of higher learning and performing linguistic and sentiment analysis to determine student perceptions in order to generate and rank those institutions based upon satisfaction and perception.

Link to faculty here

From what perspective this the program taught?

It is from a data analytical perspective, which is computational in nature and cannot be undertaken purely from a GUI-toolset perspective, so it requires students have some comfort with programming. It is statistical, but also includes machine learning, data storage and retrieval, data veracity, etc.

Does the program emphasize using data to make projections and draw conclusions or is the focus on the database design and creation?

It includes building models from data that allow us to make predictions. The base option also focuses on how to design and build data analytics systems and that includes database design and implementation. The business analytics option does not include those aspects and instead focuses on predictive and prescriptive analytics in business contexts.

Are exams proctored?

No. We will not ask you to take a proctored exam. You will mostly be doing projects, answering problems, posting to discussion boards, doing group work, etc. There are few timed exams, none of which ask you to go take them anywhere special or with a proctor.

Can you explain the difference between the base program and the option?

The base program concerns the building/design of analytic systems for any domain, focusing on the technologies whereas the business analytics option is focused on applications of data analytics to business. You’ll learn about data processing, data storage at large and ultra-large scales, etc. so use lots of different platforms such as MongoDB, Cassandra, Hadoop, Pig). In the business analytics option, students will learn to analyze large data sets for market need and to make data-driven business decisions. This track is aimed at business analysts, data scientists and analytic system designers.

If you want to learn how to build analytics systems (infrastructure/technology) and understand the technologies and platforms used in big data systems, go with the core program. Do the option if you want to know how to apply analytics in business contexts—such as buy side, supply side, and business strategy. The level of statistics is the same in each (since that is really covered in the core courses), but there is a lot more technology covered in the base program. Please let me know if you have any other questions.

I am finishing the Applied Statistics certificate. How do these programs relate?

While there is clearly overlap between the applied stat and data analytics, the primary difference between the Applied Stat masters and the option in business analytics is context. Applied Stat is focused on statistical techniques that can be applied generally to analytical problems, whereas the BA option is focused on business issues, revealing how, why and when data analytics can be used to improve decision making, forecasting, etc. Furthermore, the applied stat program is focused on statistics, whereas the data analytics program is focused on leveraging big data, and statistical approaches are just one aspect of that. The maximum number of credits you may transfer over is 12: STAT 500 and 9 credits of electives.

What tools are used?

In the program, students are exposed to: R, Python, SQL, MapReduce, Hadoop, Pig, Hive, Spark, Kafka, Weka, Cassandra-column, and MongoDB-document. Tableau and Gephi are tools used in DAAN 871. There may be other tools and technologies too.

R is used in SWENG 545 and IE 575 along with machine learning packages Weka and Knime.

Of the three BAN courses in the BAN option, two currently use R (BAN 830 and BAN 840).  In BAN 830, R is introduced for data manipulation, data visualization, and statistical analysis.  In BAN 840, R is used throughout the course where other programs like Excel would prove insufficient due to data volume capabilities or a lack of needed functionality.  Other software applications, like Tableau and GAMS, may also be used throughout the courses.

What’s the difference between applied stat and data analytics?  
Statistics play an important, fundamental, role in data analytics, which is why the Statistics Department is a partner in the Data Analytics program and several of their courses are included in it. The data analytics program offers two paths of focus – the technology, platforms, and tools of data analytics (retrieving, storing, indexing, and mining data) and the application of analytics to business problems (strategy, marketing, and supply chain). None of these topics are part of the applied stat program.

What are the classes like?

The coursework is entirely asynchronous. Lesson pages structure the week’s activities, readings, discussions, and assignments. You can expect math problems, short videos, quizzes, and technical assignments due weekly.

What is the cost for textbooks?

The cost is $150-$200 per class.

What are the objectives of the data analytics foundational course?

The course objective is to provide students with a basic background in predictive analytics and to be able to apply these techniques to large data sets. After completing this course, students should be able to build selected predictive analytic models including regression, classification and clustering using the R statistical software system.

Can I take a class or two to see if I want to do the degree program?

No, we do not offer non-degree enrollment in this program.

Is work experience required?

No, but it is beneficial.

Will my diploma say ‘online’ or look different than others at PSU?

No, World Campus does not appear on the transcript. The school issuing the degree is Penn State University—School of Graduate Professional Studies. Students who take classes online have the same looking transcript and diplomas as those who are in residence.

How long do I have to complete the degree?

Eight years.

I am interested in healthcare informatics, do you have that?

We do not offer courses in healthcare informatics currently, but it is an area we plan to expand into.

How does this program compare to other data analytics programs offered?

Most programs concentrate on business analytics, whereas ours includes an option in that domain, but also a more general option that prepares students to design and develop data analytics systems for any domain.

Will I be able to see a schedule of the next few semesters so I can schedule my classes accordingly?

Yes

Is this s STEM program?  
Yes, even the business analytics option. It is a STEM degree. Its classification code is 30.3001 Computational Science.

Does the Business Analytics option have any separate qualifications to do? How do I declare interest in that, apply separately?

It is a separate application. Reviewed by Business Analytics faculty, but the same basic admission criteria. In the business analytics option, students will learn to analyze large data sets for market need and to make data-driven business decisions. This track is aimed at business analysts, data scientists and analytic system designers.

I have an English degree but want to get into data analytics. How can I prepare myself to apply?

Take courses in algebra, probability and statistics, and some courses or experience in using databases (Access, for example).

I have transfer credits from another school, will you evaluate them?

Yes. The most you can transfer in is 9 credits, they must be within the last 5 years, you must have received a grade of B or better, and the course must be substantially similar to courses in the program (a course syllabus is usually required).

Do I need to take the classes in a particular order?

No pre-set order to the degree program courses, but we recommend starting with STAT 500, which is offered every semester). Courses are 14 weeks long. Smeal’s certificate/option courses do need to be taken in sequence, however.

How much time per week do I need to dedicate to this?

We estimate that it takes 8-10 hours per week per course for success.

Are there opportunities to work with data analytics software in any of the courses?

Yes. Most courses will use R, but some may use other packages such as SPSS and SAS.

With regards to SAS, are students exposed to statistical tools only or are they also exposed to the administration of a SAS Server?

The use of the tools only.

Is Machine Learning covered?

Yes, in IE 575, SWENG 545. You are learning about Naïve Bayes, ANN, neural nets, swarms, GAs, etc.).

Supervised learning is part SWENG 545 (so I suggest he not take STAT 508, which focuses on statistical data mining) and unsupervised learning is part of IE 575 – both core courses.

DAAN 862 includes ML and supervised learning with SciKit-Learn and natural language processing. A deep learning elective is currently under development for online delivery and should be on the schedule in late 2020.

What does the data visualization class cover?  
We like to stick to mostly open source / academic tools that are freely available. Thus, we will mostly use Sci2 <https://sci2.cns.iu.edu/user/index.php> and the visualization capabilities of R (<https://cran.r-project.org/>). We mostly use Tableau, with some Gephi. Aspects of visualization are present in other courses, such as IE 575 and INSC 846, but it is concentrated in DAAN 871, which is definitely more than an intro course.

How different is this course from different Data Science Certifications provided by Cloudera. EMC or any other organizations?

It’s a degree, not training or certification. As a university we separate education from training/certification.

Is the curriculum math intensive, writing intensive, or more hands on projects?

All of the above. Several of the courses are mathematical and analytical, while others are reading and writing intensive. There are several hands-on projects, including the capstone.

Does one gets hands on ability to solve real-time problems?

Yes, lots of hands-on, experiential learning throughout curriculum and a capstone project.

What is the capstone like?

While each capstone course focuses on problems relevant to their specific domains, they all provide students with an opportunity to apply their knowledge of the theories, methods, processes, and tools of data analytics, learned throughout their program, in a culminating and summative experience. The choice of project topic and exact form will be mutually determined by the instructor and each student. A written paper based on the applied project is required and must contain project description, analysis, and interpretation of its findings. The capstone is different for the base program and the business analytics option. In the base program students are provided a real dataset that they must load, clean, and pre-process before determining the appropriate analytics for the problem at hand and the nature of the data. Conduct the analysis and report upon their findings. In the business analytics option the focus is on the application of analytics to business decision-making with less emphasis on data management and infrastructure. It is a group project.

What kind of job titles might I apply for with this degree?

Graduates can pursue jobs as data modeler, data architect, Extraction, Transformation, Loading (ETL) developer, business intelligence (BI) developer, data warehouse developer, data analyst, systems analyst, financial analyst, etc.

What are the options for starting with the certificate and transitioning to the Master’s degree? Does the application process need to be restarted or is there a transfer option?

The Graduate Certificate in Business Analytics consists of three courses from the Data Analytics – Business Analytics option. You can begin by pursuing the certificate, and then choose to continue and apply to the master’s degree program. If you apply and are accepted to the master’s degree program, you can apply the 9 credits from the certificate toward the degree.

If I complete the 9 required BAN certificate classes while in the master’s program, can you mail me a certificate?

Yes, you have to apply for the certificate in order to receive it but can be done after the fact.

Will you accept F2 visa type in the US?

F-2 visa types are not eligible for graduate study in the face-to-face format. However, visa status is not an issue in online programs because you will not be requesting an I-20 from Penn State University.

If I fell short of a 3.0 GPA in college, but have a real-world experience, is there leeway in application consideration?

Students with less-than-ideal undergraduate GPAs can demonstrate their academic potential with high standardized test scores.

I earned my bachelor’s degree in India, where the grading is by percent of marks. How will that be converted to GPA?

The University admissions staff are well-accustomed to translating international degree transcripts into GPAs and the admissions committees are familiar with international education schemes. No-one should worry that we cannot interpret their academic performance from their respective transcripts. If we have any issues, we will request third-party evaluations or translations.

What is the level of experience the student must have on tools and platforms for these analytics programs?

We do not require students have any preparation in the tools and packages used in the program, but we do recommend having familiarity with basic coding concepts and constructs (types of data, loop constructs, sequence and selection).

Can you also talk about how we access the tools? Do we purchase them or use the cloud to connect?

Most tools are open source, so you can load them yourselves, others are server-based, and we host those. We also utilize a virtual IT –lab system that appears to the user as an environment on their local machine.

Is there a virtual lab?

The university does have virtual laboratories at our disposal. The use of these services, as well as server-side applications, depends on the course and tool. If the volume of data transfer is such that the tools cannot perform the necessary tasks in a timely fashion, students will need to install software locally. We do specify minimum technical requirements for student computers for the program at the bottom of this page: <http://www.worldcampus.psu.edu/degrees-and-certificates/data-analytics-base/apply>

Are there courses touching on project management or business process improvement?

In the BAN option the courses address analytics in a business context, including the value it can offer and the strategic advantage it presents. We offer an elective in project management, but it is a course about project management in the technology/engineering context rather than specifically in a data science context.

Do the programs include projects solving real problems of companies during the course of the program?

We do not currently partner with companies on problem sets – that presents a lot of issues regarding very large datasets, but we do use real data sets (generally from federal agencies such as NIH).

The program was designed for working professionals; therefore it is anticipated to take about 2 to 3 years to complete. Is there any way the program can be completed under one year?

Not currently. The earliest completion would be 4 or 5 semesters, depending on when you begin.

Does the program provide training in some of the tools you mentioned, such as R, Python, Hadoop, etc.?

Not training, no. Students will learn how to use the tools and platforms to complete their assignments, but the focus is not on tools training.

Python is touched upon in DAAN 822 and a full course in it is DAAN 862.

Do we need to travel to Penn State campus any time during the program?

This program does not require travel to Penn State at any time, although you are encouraged to participate in the graduation ceremony and World Campus celebration.

Are classes provided on demand, live or both?

World Campus courses are asynchronous — meaning you don't have to attend online classes at a specific time. Our students use a combination of self-study and peer-to-peer interaction over an online learning network to facilitate instruction. You'll complete weekly assignments, readings, discussions, and occasional group work, as well as exams and other activities designed to enhance learning outcomes, all at times that are most convenient to you.

What LMS do you use?

PSU and World Campus recently transitioned to Canvas, a learning management system (LMS) that offers students and faculty new ways to manage, navigate and access their courses online. The LMS offers enhanced mobile features, grading capabilities, customizable email and text message notifications, and integration with such third-party learning applications as [Turnitin](http://turnitin.psu.edu/) and [VoiceThread](http://voicethread.psu.edu/).

Will the certificate show that the degree is "online"?

Your diploma will state that The Pennsylvania State University grants you the degree of Master of Professional Studies in Data Analytics. Neither the diploma nor the transcript will differentiate the mode (online or otherwise) in which you completed the courses or degree.

What are your thoughts about this career as a mid-life career change?

It is ideally suited to those seeking a mid-life career change. The tools and techniques in our program could do not implemented at large scales only a few years ago, so very few professionals have them now. The advantage for those in mid-career is that they already have the contextual understanding - whether they are in finance, marketing, healthcare, pharma, supply-chain, etc. they understand the business problem already and now can learn the science and techniques that can be leveraged.

In traditional brick and mortar schools there are resources for tutoring, faculty office hours, open lab help, etc. In this format, how do students get help if they need assistance?

Faculty are very accessible through phone and email and respond to students in a very timely manner. To help you complete your work when it is most convenient for you, our technical support and academic advising staff are available through extended work hours. And they are easy to access by email, phone, and instant messaging.

Will an advisor/mentor be assigned?

We use an advising helpline that students can reach by email or phone to answer any questions or concerns they have or to set up specific counselling appointments.

What makes Penn State's data analytics program unique?

It is comprehensive, tailorable, and delivered by faculty and units that have the highest reputation and ranking in their respective fields using state-of-the-art platforms and evidence- and research-based learning design.

What is marking analytics?

It is understanding the customer (so things like text mining/sentiment analysis in social media; buying habits; loyalty program data analysis; customer interests) as well as things like analysis of marketing efforts (Google ads, click throughs, SEO).

Why do marketing analytics?

One of our differentiators is that we cover MAN from an outside-in approach - meaning teach the topics from the marketing challenge first and align data/stats to them. Other programs teach from the stats (e.g. they have a course called “Regression 101”) and students are expected to figure out how to fit them to the business. We want students to leave the marketing analytics track with a toolkit of analytics they can easily apply to common marketing challenges they’ll most certainly encounter in the workplace.

When are BAN classes offered?

|  |  |
| --- | --- |
| **Course** | **When Offered** |
| BAN 830 | Fall, Spring |
| BAN 840 | Spring, Summer |
| BAN 550/850 | Fall, Summer |

Great Valley offers the MS in residence. While the Master in Data Analytics is intended to provide professionally-oriented technical education, there is a segment of the student population interested in pursuing a research-oriented education focused on advancing the techniques and tools of data analytics as well as the application of new and existing techniques to previously unexplored domains and situations. It is the objective of the Master of Science in Data Analytics to prepare those students for entry into doctoral programs in data analytics.

What are alumni doing?

The most common job titles of our alumni are Manager, Engineer, and Consultant. Three companies have more alumni working in than others: Amazon, Comcast, and Boeing.