

Assignment 7 - Project Report

Portland State University Student Graduation Rate Analysis

Course Title: Data Warehousing **Course Number**: ETM 538 **Instructor:** Dr. Mike Freiling, Prof. Daniel Sagalowicz

Term: Winter Year: 2017 Author(s): Chengjun Li, Obinna Mbagwu Sejoong Kim, Shivani Purwar, Surekha Chanamolu Date: 03/08/2017

ETM OFFICE USE ONLY

Report No.: Type: Student Project Note:

Table of Contents

Portland State University Background2	
Project Interview2	2
Concepts the PSU OIT using to analyze the data2	
Data Provided by PSU4	-5
Project Requirements3	-15
Appendix	16

1.0 Portland State University Background:

Portland State University (PSU) is a public, non-profit, coeducational research university located in the Southwest University District of downtown Portland, Oregon, United States. It was founded in 1946 as a post-secondary educational institution for World War I Veterans. It evolved into a four-year college over the following two decades, and was granted university status in 1969. It is the only public urban university in the state of Oregon that is located in a major metropolitan city, and is governed by a board of trustees.

As Oregon's most affordable public research university, Portland State University offer tremendous opportunity to 27,000 students from all backgrounds. Their mission is "let knowledge serve the city" reflects their dedication to finding innovative, sustainable solutions to local and global problems.

Enrollment 2016-2017 (Fall 2016) Total: 27,229 Undergraduate: 21,633 Graduate: 5,596 Women: 14,856 Men: 12,373 Full time: 62% Part time: 38% Student to Faculty Ratio: 17 to 1

2.0 Project Interview

Interviewed the PSU OIT Business Intelligence Team Manager on Feb.10th 2017 to identify business needs of BI team of PSU and the datasets from their data warehouse that can be used to answer identified business questions needed to achieve the key business goals.

2.1 Key Business Goals of the PSU OIT Business Intelligence Team:

The key business goal of this project is to improve the graduation rates of the PSU students.

3.0 Key Business Questions:

The key business questions this project would like to address are:

- 1. Who has graduated from PSU?
- 2. Who is still attending PSU?
- 3. Of the stop outs from PSU:
 - a. Who has gone on to other schools?

- i. Is there a pattern in the data of certain majors moving on to programs at another school?
- b. Who has not gone to other schools? (Who has stopped out?)
 - i. The students who have not moved onto other schools, do they need assistance from advisors to complete their degree or not?

4.0 Concepts the PSU OIT using to analyze the data:

Currently they do not have any analysis data that they can provide us.

4.0 Data Provided by PSU:

1. random_clearinghouse.csv

Contains first time freshmen (PSU freshmen students who came to PSU straight from high school) who began in the same term (cohort) and the colleges they have attended according to the National Student Clearinghouse. The identifiers have been randomized, but match between files.

2. random_psu_degrees.csv

Degrees awarded to the students from PSU.

3. random_psu_terms_majors.csv First and last terms attended for students along with their most recent major.

5.0 Project Requirements:

5.1 Key Attributes to use:

The key attributes to determine the key business questions (Who has graduated from PSU? Who is still attending PSU? Who has stopped out? Who has gone on to other schools? Is there a pattern in the data of certain majors moving on to programs at another school? Who has not gone to other schools?) in the random_clearinghouse.csv, randon_psu_degrees.csv and random_psu_terms_majors.csv tables are highlighted in blue as shown in tables 1, 2 and 3 below.

random_clearinghouse.csv		
Attribute	Data Type	Description
FIRST_NAME	Short Text	Randomized first name for each student
LAST_NAME	Short Text	Randomized last name for each student
COHORT_TERM	Number	Group of students who started at PSU in the same term
RANDOM_UID	Number	Randomized identifier for the student
REC_FOUND_IND	Short Text	
COLLEGE	Number	1026-42118, The code of each school
COLLEGE_BRANCH	Number	0,1,2,3,4,5,6,9,10,15,20,40,47
COLLEGE_DESC	Short Text	The name of schools
COLLEGE_STATE	Short Text	The location of schoos
COLLEGE_2_4_YEAR	Short Text	2,4,L
PUBLIC_PRIVATE	Short Text	Private or public school
ENROLL_BEGIN_DATE	Date/Time	2007-2016
ENROLL_END_DATE	Date/Time	2007-2016
		Codes for this have 01 for Winter, 02 for Spring, 03 for Summer and 04
ENROLL_TERM	Number	for Fall term. For example, 201604 is the code Fall 2016
ENROLL_STATUS	Short Text	A,F
		Decreased, Full-time, Half-time, Leave of absence, Less than half-time,
ENROLL_STATUS_DESC	Short Text	Three-quarter time, Withdrawn
CLASS_LEVEL	Short Text	A,B,C,D,F,G,J,L,M,N,P,R,S,T
		Associates, Bachelors(Undergraduate), Certificate(Undergraduate), Doctoral(Graduate), First Professional(Graduate), Freshman(Undergraduate), Junior(Undergraduate), Masters(Graduate), Post Baccalaureate Certificate, Postdoctorate(Graduate), Senior(Undergraduate), Sophomore(Undergraduate),
CLASS_LEVEL_DESC	Short Text	Unspecified(Graduate/Professional), Unspecified(Undergradutae)
GRADUATE_UNDERGRADUATE	Short Text	Graduate, Undergraduate
ENROLL_MAJOR1	Short Text	The name of majors
		The Classification of Instructional Programs (CIP) provides a taxonomic
		scheme that supports the accurate tracking and reporting of fields Of
ENROLL_MAJOR1_CIPC	Short Text	study and program completions activity.
ENROLL_MAJOR2	Short Text	
ENROLL_MAJOR2_CIPC	Short Text	
GRADUATED_IND	Short Text	Student graduation status - N - No, Yes -Y
GRADUATION_DATE	Short Text	2007-2016
DEGREE_DESC	Short Text	
DEGREE_MAJOR1	Short Text	The type of degrees
DEGREE_MAJOR1_CIPC	Short Text	
DEGREE_MAJOR2	Short Text	
DEGREE_MAJOR2_CIPC	Short Text	
DEGREE_MAJOR3	Short Text	
DEGREE_MAJOR3_CIPC	Short Text	
DEGREE_MAJOR4	Short Text	
DEGREE MAJOR4 CIPC	Short Text	

Table 1. random_clearinghouse

random_psu_degrees.csv			
Attribute	Data type	Range (number only)	Description
RANDOM_UID	Number	1559903-135586638	ID number
			That is the term they graduated. Eg. The field says 201604 then they
ACADEMIC_PERIOD_GRADUATION	Number	201001-201701	graduated in Fall 2016 term.
DEGREE	Short Text		To determine which degree granted by PSU.
DEGREE_DESC	Short Text		Full name of degrees abbreviation
MAJOR	Short Text		The main major when the stududent graduated
MAJOR_DESC	Short Text		Full name of majors abbreviation
			Student level are determined based upon a student's academic
STUDENT_LEVEL	Short Text		program when they graduated.
STUDENT_LEVEL_DESC	Short Text		Here are 2 levels: UG-undergraduate GA-graduated
			The code of college, each number indicate a single college. Total 26
COLLEGE	Number	from 1 to 26	codes
COLLEGE_DESC	Short Text		determin which number to which college
SECOND_MAJOR	Short Text		The second major when the stududent graduated
SECOND_MAJOR_DESC	Short Text		Full name of majors abbreviation
THIRD_MAJOR	Short Text		The third major when the stududent graduated
THIRD_MAJOR_DESC	Short Text		Full name of majors abbreviation
FOURTH_MAJOR	Short Text		The fourth major when the stududent graduated
FOURTH_MAJOR_DESC	Short Text		Full name of majors abbreviation
FIFTH_MAJOR	Short Text		The fifth major when the stududent graduated
FIFTH_MAJOR_DESC	Short Text		Full name of majors abbreviation
SIXTH_MAJOR	Short Text		The sixth major when the stududent graduated
SIXTH_MAJOR_DESC	Short Text		Full name of majors abbreviation

Table 2. random_psu_degrees

random_psu_terms_majors.csv			
Attribute	Data type	Range (number only)	Description
RANDOM_UID	Number	215401-152791898	Random Student ID
			The first term student enrolled at PSU. Eg. 201404 is the fall
PSU_FIRST_ACADEMIC_PERIOD	Number	199504-201604	term at 2014.
			The last term student enrolled at PSU. Eg. 201401 is the winter
PSU_LAST_ACADEMIC_PERIOD	Number	200704-201701	term at 2014.
			Student level are determined based upon a student's academic
			program at the start of the term. Here are 4 levels:
			UG-undergraduate PB-post-baccalaureate
STUDENT_LEVEL_PRI1	Short Text		GA-graduated NA- Undeclared/Not Applicable
MAJOR1_PRI1	Short Text		Primary major of student at the start of the term
MAJOR1_DESC_PRI1	Short Text		Short description of primary majors.
MAJOR2_PRI1	Short Text		The second major of student at the start of the term.
MAJOR2_DESC_PRI1	Short Text		Short deccription of second marjors.
MAJOR3_PRI1	Short Text		Here it is the third major of student at the start of the trem.
MAJOR3_DESC_PRI1	Short Text		Short deccription of third marjors.
MAJOR4_PRI1	Short Text		Here it is the forth major of student at the start of the trem.
MAJOR4_DESC_PRI1	Short Text		Short deccription of fourth marjors.
MAJOR1_PRI2	Short Text		The main major which after students changed
MAJOR1_DESC_PRI2	Short Text		Short deccription of main marjors after students changed.
MAJOR2_PRI2	Short Text		The second major which after students changed
MAJOR2_DESC_PRI2	Short Text		Short deccription of second marjors after students changed.
MAJOR3_PRI2	Short Text		The third major which after students changed
MAJOR3_DESC_PRI2	Short Text		Short deccription of third marjors after students changed.
MAJOR4_PRI2	Short Text		The fourth major which after students changed
MAJOR4_DESC_PRI2	Short Text		Short description of fourth marjor after students changed.

Table 3. random_psu_terms_majors

5.2 Any bucketing you plan to use for key attributes.

After reviewing our dataset tables we discovered that students admitted in the same term in any given year is being classified in that same year and term cohort. We decided to bucket each student in their cohort terms as shown in the table below.

COHORT TERM	Description	Bucket
200704	FALL 2007	Α
200804	FALL 2008	В
200904	FALL 2009	С
201004	FALL 2010	D
201104	FALL 2011	E
201204	FALL 2012	F
201304	FALL 2013	G
201404	FALL 2014	Н
201504	FALL 2015	

Table 4. cohort bucketing

5.3 Algorithms you think are worth trying. (Only algorithms that were presented in the class are allowed)

1R Algorithm: By combining PSU datasets of Random_psu_term_major containing all the psu cohort and the random_degree_psu containing the students that have graduated from PSU, we developed a 1R to determine which students have graduated in a selected cohort. If the random id in the graduation table is equal to the random id of a student in a cohort then graduated equal 'YES' else graduated equal 'NO'.

Bayesian Algorithm (Naive Bayes): Using the data of graduated (graduated =YES) and not graduated (graduated =NO) students we can determine the probability of a student graduating from PSU in a given cohort, degree, major, student level and college.

5.4 Evaluation criteria.

1. Students who have graduated from PSU -

Using 1R algorithm if a student id in a cohort in PSU graduated students table(random_psu_degrees) exists in the psu all cohorts table(random_psu_terms_majors) then the student's graduation status equal 'YES' else if the student id is not present in the PSU graduation table then the student's graduation status is equal 'NO'. The NO status forms the category of students used in further analysis required to answer other business questions. psu_terms_majors-GradStatus- No table contain the data of students who have never graduated from PSU.

PSU COHORT	Graduation = No	Graduation = Yes	Total no. of students in the cohort	Error Rate - Graduation - No
200704	777	678	1455	0.534020619
200804	881	691	1572	0.56043257
200904	860	682	1542	0.55771725
201004	733	604	1337	0.548242334
201104	735	500	1235	0.5951417
201204	1003	283	1286	0.779937792
201304	1479	23	1502	0.984687084
201404	1552	1	1553	0.999356085
201504	1691	0	1691	1
Grand Total	9711	3462	13173	0.737189706

Table 5. Students who have graduated from PSU



Fig 1.Graph of PSU student Graduation

2. Students who are not graduated in the Cohort and still attending PSU -By applying 1R rule to PSU_LAST_ACADEMIC_PERIOD attribute in random_psu_term_majors tables the student RANDOM_UID is compared to RANDOM_UID for the students not graduated in the clearing house data to determine which student ids in PSU are among the not graduated RANDOM_UID in the clearing house and with PSU_LAST_ACADEMIC_PERIOD = 201701 (winter 2017). The currently enrolled but not graduated PSU students grouped by their cohorts is as shown in table 7 and figure 3 below.

	Total No. of Students in	Students Still attending
Cohort Term	Cohort	PSU
200704	1455	23
200804	1572	28
200904	1542	45
201004	1337	68
201104	1235	121
201204	1286	364
201304	1502	756
201404	1553	879
201504	1691	1101

Table 6. Students are still attending PSU



Fig 2. Graph of students who are still attending PSU

3. Students who are not graduated in the Cohort and transferred to other schools By applying 1R rule to random_clearinghouse where COLLEGE_DESC attribute not equal to PSU we extracted all the students ids (RANDOM_UID) enrolled in other colleges and compared these RANDOM_UIDs to the RANDOM_UIDs of random_psu_terms_majors to determine which of the not graduated students in each cohort have transferred to other institutions. The outcome is as shown in table 8 and figure 4.

	Total No. of PSU	
	Students in	Total Tranferred
Cohort Term	Cohort	Students
200704	1455	307
200804	1572	356
200904	1542	345
201004	1337	311
201104	1235	304
201204	1286	400
201304	1502	466
201404	1553	380
201504	1691	139

Table 8: Transferred Students in each Cohort



Fig 4. Graph of the Transferred Students in each Cohort

4 Students who have stopped out -

Comparing the clearinghouse RANDOM_UID to the RANDOM_UID of psu_terms_majors we determined which student is not currently enrolled where attribute the latest EROLL_TERM in the clearinghouse attribute is less than 201701 (Winter 2017).

	Total No. of Students in	Total No. of Stopped Students in the Cohort -
Cohort Term	Cohort	Never Graduated
200704	1455	521
200804	1572	609
200904	1542	590
201004	1337	532
201104	1235	529
201204	1286	600
201304	1502	708
201404	1553	669
201504	1691	590



Fig.2 Graph of Students Stopped - Never graduated

5. Transferred Students by Major -

Using the outcome of step 3 we determined the transfers in each major by filters on MAJORS attributes.

PSU Major	Total No. Transferred Students
Advertising Management Minor	1
Anthropology	23
Applied Health & Fitness	2
Applied Linguistics	13
Architecture	54
Art	1
Art History	3
Art Practices (BFA)	7
Art: Art History	2
Art: Art Practices	34
Art: Drawing/Pntng/Prntmkng	25
Art: Graphic Design	69
Art: Sculpture	4
Arts and Letters	16
Arts Studies	6
Biology	141
Biology: Micro/Molecular	18
Biology: Organismal	13
Biomedical Informatics 3+2	7
Black Studies	1
Bus Ad: General	1
Bus Ad: Real Estate Finance	7
Bus Ad: Supply & Logistic Mgmt	19
Bus Adm: Mgmt & Leadership	68
Business Adm: Accounting	58
Business Adm: Advertising Mgmt	32
Business Adm: General Mgmt	46
Business Adm: Human Res Mgmt	29
Business Adm: Info Systems	3
Business Adm: Marketing	96
Business Admin: Finance	39
Business Administration	2
Chemistry	26
Chemistry: Biochemistry	18
Child & Family Studies	34
Civil Engineering	41
Communication	1
Communication Studies	68
Community Development	8
Computer Engineering	31
Computer Science	93

PSU Major	Total No. Transferred Students
Conflict Resolution	2
Creative Writing	1
Criminology & Criminal Justice	73
Earth Science	2
Economics	29
Education (MEd)	1
Electrical Engineering	45
Elem Ed	1
Elementary Education Minor	5
English	90
Environmental Engineering (BS)	10
Environmental Sciences	26
Environmental Studies	37
Exploratory/Undeclared	326
Film	52
Foreign Languages: Arabic	1
Foreign Languages: Chinese	2
Foreign Languages: French	6
Foreign Languages: Japanese	6
Foreign Languages: Russian	2
Foreign Languages: Spanish	8
Geography	7
Geology	7
Health Ed: Health Sciences	2
Health Studies	4
Health Studies: Aging Services	3
Health Studies: Community	39
Health Studies: Health Science	166
Health Studies: School Health	6
History	34
HIth Studies: Phys Actvty-Exerc	46
Intl & Glbl St: European Stds	1
Intl & Glbl St: Intl Develpmnt	2
Intl & Glbl St: Middle East	1
Intl Studies	3
Intl Studies: African Studies	4
Intl Studies: East Asian	6
Intl Studies: European Studies	11
Intl Studies: Intl Development	4
Intl Studies: Latin America	6
Intl Studies: Middle East	5

PSU Major	Total No. Transferred Students
Liberal Studies	55
Mathematics	31
Mechanical Engineering	66
Music (BM)	4
Music: Composition (BM)	2
Music: Education (BM)	3
Music: General	18
Music: Jazz (BM)	4
Music: Performance (BA/BS)	1
Music: Performance (BM)	11
Music: Voice (BM)	8
Philosophy	17
Physics	11
Physics: Biomedical	1
Poli Sci: Intl Development	1
Political Science	41
Pre-Law	1
Pre-Medicine	1
Pre-Nursing	1
Professional Writing Minor	1
Psychology	168
Science	52
Secondary Education Minor	1
Social Science	27
Social Work - Bachelor's	27
Sociology	37
Special Education Minor	1
Speech & Hearing Sciences	17
Theater Arts	33
Undeclared/Not Applicable	88
Urban Studies (MUS)	1
Visual Art	1
Womens Studies	7
World Language: Arabic	2
World Language: Chinese	4
World Language: French	2
World Language: German	2
World Language: Japanese	5
World Language: Russian	4
World Language: Spanish	8
Writing Minor	1



6.0 Appendix:

Definition of terms

Academic Period: Term (ie Fall 2016 term, Winter 2017 term etc). Codes for this have 01 for Winter, 02 for Spring, 03 for Summer and 04 for Fall term. For example 201604 is the code Fall 2016. Cohort: Group of students who started at PSU in the same term

Random UID: Randomized identifier for the student

Stop Out: Student has not attended PSU for 1 or more terms.

Table Specs

CREATE TABLE "RANDOM_CLEARINGHOUSE"

"FIRST_NAME" VARCHAR2(255 BYTE), ("LAST_NAME" VARCHAR2(255 BYTE), "COHORT_TERM" VARCHAR2(24 BYTE) "RANDOM UID" VARCHAR2(1020 BYTE), "REC FOUND IND" VARCHAR2(255 BYTE), "COLLEGE" VARCHAR2(24 BYTE), "COLLEGE_BRANCH" VARCHAR2(8 BYTE), "COLLEGE_DESC" VARCHAR2(255 BYTE), "COLLEGE_STATE" VARCHAR2(255 BYTE), "COLLEGE_2_4_YEAR" VARCHAR2(255 BYTE), "PUBLIC PRIVATE" VARCHAR2(255 BYTE), "ENROLL BEGIN DATE" DATE, "ENROLL END DATE" DATE, "ENROLL_TERM" VARCHAR2(63 BYTE), "ENROLL_STATUS" VARCHAR2(255 BYTE), "ENROLL_STATUS_DESC" VARCHAR2(19 BYTE), "CLASS_LEVEL" VARCHAR2(255 BYTE), "CLASS_LEVEL_DESC" VARCHAR2(35 BYTE), "GRADUATE_UNDERGRADUATE" VARCHAR2(13 BYTE), "ENROLL_MAJOR1" VARCHAR2(255 BYTE), "ENROLL_MAJOR1_CIPC" VARCHAR2(63 BYTE), "ENROLL_MAJOR2" VARCHAR2(255 BYTE), "ENROLL MAJOR2 CIPC" VARCHAR2(63 BYTE), "GRADUATED_IND" VARCHAR2(255 BYTE), "GRADUATION_DATE" DATE,

"DEGREE_DESC" VARCHAR2(255 BYTE), "DEGREE_MAJOR1" VARCHAR2(255 BYTE), "DEGREE_MAJOR1_CIPC" VARCHAR2(63 BYTE), "DEGREE_MAJOR2" VARCHAR2(255 BYTE), "DEGREE_MAJOR2_CIPC" VARCHAR2(63 BYTE), "DEGREE_MAJOR3" VARCHAR2(255 BYTE), "DEGREE_MAJOR3_CIPC" VARCHAR2(63 BYTE), "DEGREE_MAJOR4" VARCHAR2(255 BYTE), "DEGREE_MAJOR4" VARCHAR2(255 BYTE),

);

CREATE TABLE "RANDOM_PSU_DEGREES"

("RANDOM_UID" NUMBER, "ACADEMIC_PERIOD_GRADUATION" VARCHAR2(63 CHAR), "DEGREE" VARCHAR2(63 CHAR), "DEGREE_DESC" VARCHAR2(255 CHAR), "MAJOR" VARCHAR2(63 CHAR), "MAJOR_DESC" VARCHAR2(255 CHAR), "STUDENT LEVEL" VARCHAR2(63 CHAR), "STUDENT LEVEL DESC" VARCHAR2(255 CHAR), "COLLEGE" VARCHAR2(63 CHAR), "COLLEGE_DESC" VARCHAR2(255 CHAR), "SEC_MAJOR" VARCHAR2(63 CHAR), "SEC_MAJOR_DESC" VARCHAR2(255 CHAR), "THIRD MAJOR" VARCHAR2(63 CHAR), "THIRD MAJOR DESC" VARCHAR2(255 CHAR), "FOURTH MAJOR" VARCHAR2(63 CHAR), "FOURTH_MAJOR_DESC" VARCHAR2(255 CHAR) "FIFTH_MAJOR" VARCHAR2(63 CHAR), "FIFTH_MAJOR_DESC" VARCHAR2(255 CHAR) "SIXTH_MAJOR" VARCHAR2(63 CHAR), "SIXTH_MAJOR_DESC" VARCHAR2(255 CHAR));

CREATE TABLE "RANDOM_PSU_TERMS_MAJORS"

("RANDOM_UID" NUMBER,
"PSU_FIRST_ACADEMIC_PERIOD" VARCHAR2(63 CHAR),
"PSU_LAST_ACADEMIC_PERIOD" VARCHAR2(63 CHAR),

"STUDENT_LEVEL_PRI1" VARCHAR2(63 CHAR), "MAJOR1_PRI1" VARCHAR2(63 CHAR), "MAJOR1_DESC_PRI1" VARCHAR2(255 CHAR), "MAJOR2_PRI1" VARCHAR2(63 CHAR), "MAJOR2_DESC_PRI1" VARCHAR2(255 CHAR), "MAJOR3_PRI1" VARCHAR2(63 CHAR), "MAJOR3_DESC_PRI1" VARCHAR2(255 CHAR), "MAJOR4_PRI1" VARCHAR2(63 CHAR), "MAJOR4_DESC_PRI1" VARCHAR2(255 CHAR) "MAJOR1_PRI2" VARCHAR2(63 CHAR), "MAJOR1_DESC_PRI2" VARCHAR2(255 CHAR) "MAJOR2_PRI2" VARCHAR2(63 CHAR), "MAJOR2_DESC_PRI2" VARCHAR2(255 CHAR), "MAJOR3_PRI2" VARCHAR2(63 CHAR), "MAJOR3_DESC_PRI2" VARCHAR2(255 CHAR), "MAJOR4_PRI2" VARCHAR2(63 CHAR), "MAJOR4_DESC_PRI2" VARCHAR2(255 CHAR));