Overview:

We began this project hoping to understand what affects both short-term and long-term success of Indian students. To this end, we have completed two separate analyses: one of the UIDSE state-level data on primary school dropout, and the other on the AspiringMinds data for long-term skills and employment.

Before we begin: Teacher quality and accountability, accountability for distribution of funds, and home life/parents' attitudes are and will always be the most important factors keeping students in schools. This analysis attempts to optimize government investment in schools themselves, however these 3 factors should always be considered with the highest priority, including more collection of data so that we can address the problems with each.

Analysis of Primary School Dropout:

We analyzed the UIDSE 2014-2015 state-level data on student dropout rates to find the school-related qualities that made the biggest difference in keeping students in schools. To find statistically significant effect of each of the changes, such as school size, toilet access, etc., we used a two-tail paired t test. We constructed the ratio of [students who attended primary only] to [students who attended anything above only primary] for each state and each attribute (results included below each attribute's data in Excel spreadsheet). We then compared the statistics for all the states to the "Total Schools" averages to see if each attribute made a significant impact.

At our discretion, we also compared certain related attributes with each other to see if there was a significant difference between them. For example, we compared "Single-classroom schools" with "Number of classrooms per school".

Summarized results are as follows:

1) Government-aided schools have significantly less dropouts, so government assistance does seem to be helping although there is of course room for improvement
2) Providing a mid-day meal definitely helps keep students in school.
    a. Food prepared on premises is more effective
    b. Just providing kitchen shed increases retention
3) Small schools (<=50 students) clearly have more dropouts
    a. Single classroom schools clearly have more dropouts as compared to "Number of classrooms", corroborating this theory
4) Single teacher schools are even more prone the dropouts than small schools or single-classroom schools
5) Providing toilets, both girls and boys, drastically decrease dropout rate
6) New schools (2001 or later) have significantly higher dropout rates
7) Adding infrastructure like all-weather accessible roads, playgrounds, and a boundary wall significantly improve retention
8) Significantly more girls drop out after primary school than boys
9) Schools with computers and electricity show a vast increase in retention, however schools with both these things probably have other amenities that account for some of this retention (like toilets, food, a playground, multiple classrooms and teachers, etc.)

10) There seems to be no significant difference between schools with and without ramps, however this is probably because handicapped students are a minority and do not show up in these large averages. Ramps are important!

Analysis of Long-Term Skills:

We also analyzed the AspiringMinds data of a wide range of engineering students collected several years after graduation. For the sake of interpretability, which is critical in this application, we used a logistic regression classifier to predict whether a person would have any skill in each of the different areas of engineering. It was important to us that we predict only skill, not qualifications, wealth, job or any other external factors. This would also allow us to find highly-skilled people who were likely being overlooked.

We provided the model all data about a person's life and personality through college years, and analyzed the coefficients of the logistic regression model (Appendix A) for each type of engineering. To analyze the effectiveness of the model itself, we used the area under the ROC curve as well as training/testing accuracies, also included in the appendix (and the Jupyter notebook).

Overall trends showed that having taken a 10^{th} or 12^{th} board exam in a related field increased likelihood of having future skill in an engineering field. This means both that the exams were effective in encouraging actual skill and learning, but also that intervention to encourage diversity and minorities to take an interest in these fields needs to happen at a much younger age, beginning in primary school.

A detailed analysis of the factors most affecting (positively or negatively) a person's likelihood of having skill are included in a field-by-field analysis below:

1) Computer Science: This model had the most data as well as the highest ROC (0.82), so we can be confident in these results. Interestingly, being female was the top positively correlating factor that determined skill in Computer Science. What is indicates, other than the fact that India is producing many highly-skilled female computer scientists, is a need to encourage women and girls in the other areas of engineering. This is because our model is classifying computer scientists versus all other types of engineers, and female computer scientists are proportionally higher. Colleges in Jammu/Kashmir, Orissa, Himachal Pradesh, and Punjab will also benefit from government encouragement of computer science, as these are some of the college states negatively correlated with computer science skill.

2) Computer Programming: Colleges in Tamil Nadu, Andhra Pradesh, Jharkhand, and Telangana, in that order, are particularly associated with skill in computer programming. Women are also well represented. Colleges in Assam are negatively correlated with skill.

3) Electronics Engineering: Colleges that are negatively correlated with skill and could use help are in Haryana, Uttarakhand, Punjab, Maharashtra, and Himachal Pradesh. Women need particular encouragement here. Colleges in Tamil Nadu and Jharkhand are positively associated with skill.

4) Mechanical Engineering: Colleges in West Bengal, Haryana, and Delhi are negatively correlated, as are females.
5) Telecom Engineering: Colleges in Andhra Pradesh and Himachal Pradesh are negatively correlated.
6) Civil Engineering: Colleges in Tamil Nadu, Haryana, Orissa, Telangana, Rajasthan, and Delhi are negatively correlated, as are females. Colleges in Jharkhand and Himachal Pradesh are positively correlated.

Recommendations:

When investing in schools, the government's priority should be safe and clean toilets for both genders, as well as a fresh midday meal cooked on school premises. These are the easiest, most basic ways to boost retention rates through primary school.

The next step after these necessities is investment in electricity and computers, as well as more teachers and classrooms. It goes without saying that quality of teaching and teacher accountability is critical throughout their careers. However, even simply adding more teachers creates a statistically significant increase in retention when compared to single-teacher schools. Computer recycling programs to take old computers out of private schools and donating them to public schools could easily bridge some of the gaps in computer access. Government investment in universal internet access and electricity will empower the students who are now left out.

In terms of infrastructure, although they may cost more the following are worth investing in: a boundary wall around the school, an all-weather accessible road, and a playground. Interestingly, newer schools (established since 2001) have a much higher chance of dropouts. This warrants further investigation, and may simply be due to confounding factors like lack of food, electricity, toilets, and classrooms/teachers, but for the purposes of this analysis it indicates that instead of investing in new school buildings, the government should invest in the roads, playgrounds, and boundary walls.

The fact that more significantly more girls drop out than boys means that effort should be put forward to impress the importance of educated women both on the girls and their families. This is a longer process, but nevertheless a critical one, and is closely linked to making girls feel safe at school with toilets, food, multiple teachers, and accessible roads.

Recommendations for spending on the college level are as follows, with a focus on early intervention in the states and groups at risk of falling behind for each field of engineering:

1) Computer Science: Invest in colleges in Jammu/Kashmir, Orissa, Himachal Pradesh, and Punjab as these are the college states negatively correlated with computer science skill. Invest in encouraging girls to pursue fields of engineering outside of just computer science and computer programming.

2) Computer Programming: Focus more investment outside of only Tamil Nadu, Andhra Pradesh, Jharkhand, and Telangana, and invest more in Assam. Once again, encourage girls to pursue all kind of engineering, as they are well represented here as well.

3) Electronics Engineering: Invest in electronics education in colleges in Haryana, Uttarakhand, Punjab, Maharashtra, and Himachal Pradesh, with a particular focus on encouraging more women.
4) Mechanical Engineering: Invest in mechanical engineering programs in colleges in West Bengal, Haryana, and Delhi, with a particular focus on encouraging more women
7) Telecom Engineering: Invest in telecom engineering programs in colleges in Andhra Pradesh and Himachal Pradesh.
5) Civil Engineering: Invest in civil engineering programs in colleges in Tamil Nadu, Haryana, Orissa, Telangana, Rajasthan, and Delhi, with a particular focus on encouraging women.

Thank you for taking a look at this analysis, and we hope it was helpful!

Appendix A: Factors most contributing to success in each type of engineering

SKILL: ComputerProgramming
Train acc: 0.787191460974
Test acc: 0.783
ROC AUC: 0.57605762688

Top 10: {'12board_mp board': 0.53107448364540988, 'CollegeState_Tamil Nadu': 0.5556516410857841 3, 'CollegeState_Andhra Pradesh': 0.66860264794118196, '12board_rajasthan board': 0.4122491976074 5743, 'CollegeState_Jharkhand': 1.2591695848350495, '12board_central board of secondary education': 0.61491648165629631, 'Gender_f': 0.59401549198873094, '12board_bie': 0.60140850048698691, 'Colle geState_Telangana': 0.52319600337538463, '12board_board of intermediate education': 1.00294358535 89426}

Bottom 10: {'10board_matriculation': -0.75610992911494368, '12board_nios': -0.88302774157556996, '1 0board_tamilnadu state board': -0.8016632880331882, '10board_matric': -0.62566483559956765, '12boa rd_tamilnadu state board': -0.8016632880331882, '10board_bse,orissa': -0.68484467007666483, 'Colleg eState_Assam': -0.54307942768490569, '12board_pu board': -0.78505105600376557, '12board_chse,ori ssa': -0.68484467007666483, '10board_kseeb': -0.78226094552451808}

SKILL: ElectronicsAndSemicon
Train acc: 0.75083388926
Test acc: 0.707
ROC AUC: 0.592616288369

Top 10: {'10board_cbse board': 0.95506042327191665, '10board_board of school education harayana': 0.83693726458473183, '10board_bihar board': 1.153919600452388, '10board_state': 0.84624470542692 631, '10board_matriculation board': 0.82061557838966426, '10board_karnataka secondary education bo ard': 1.1216988974466588, '12board_jharkhand academic council': 0.80565232705717527, '12board_p u board, karnataka': 0.75569552310738841, '12board_pu board': 0.79653301176983882, '10board_kseeb': 0.91250138661201874}

Bottom 10: {'10board_rbse': -0.93107769480109648, '12board_chse': -0.90386478356696587, '12board _up board': -0.66157432836756913, '12board_hbse': -0.84071080650113206, '12board_puc': -0.7122617 4366131367, '10board_sslc': -0.63164708309988815, '10board_upboard': -0.62965823104770813, '12bo ard_upboard': -0.65259950888167773, 'CollegeState_Assam': -0.8200036447574266, '10board_mp boar d': -0.7633164595437214}

SKILL: ComputerScience
Train acc: 0.791527685123
Test acc: 0.765

ROC AUC: 0.822076612903

Top 10: {'Gender_f': 0.76969762156217292, '12board_cbse': 0.68847735057808712, 'Gender_m': 0.608
97536512880679, '12board_icse': 0.69547197111609782, '10board_icse': 0.72965680398835431, '10bo
ard_state board': 0.96875315162776965, '12board_state board': 0.77231003956826716, '12board_pseb':
0.52945413366309546, '10board_pseb': 0.52945413366309546, '10board_cbse': 0.70374513502394598
}

Bottom 10: {'years_since_grad': -0.7438205209913501, '10board_ssc': -0.75998304568899055, '10boar
d_board of secondary education': -0.23370587556916755, '12board_board of intermediate education': -0.
24141212106853754, 'CollegeState_Punjab': -0.37892752747222946, 'CollegeState_Union Territory': -0.
26372879437923202, '10board_0': -1.3155854689343374, 'CollegeState_Jammu and Kashmir': -0.34205
365239456503, 'CollegeState_Orissa': -0.23918762416866446, 'CollegeState_Himachal Pradesh': -0.240
2081068493257}

SKILL: MechanicalEngg
Train acc: 0.942294863242
Test acc: 0.94
ROC AUC: 0.660372340426

Top 10: {'12board_intermidiate': 0.76117558780109973, '10board_cbse ': 0.7955124758860953, '10boar
d_matriculation board': 0.78689615447525929, '12board_staae board': 0.84155993216958302, '12board
_state board of technical eduction panchkula': 0.77955020696712496, '12board_state broad': 0.8515349
5329656925, '10board_upboard': 1.0077633214059252, '12board_jharkhand acamedic council (ranchi)':
0.75320045211386777, '10board_jharkhand secondary examination board (ranchi)': 0.753200452113867
77, '12board_tamil nadu state board': 0.78689615447525929}

Bottom 10: {'10board_up board': -0.48881345533616616, '10board_state board': -0.5285662065246001
8, '12board_board of intermediate': -0.96682645233398024, 'CollegeState_West Bengal': -0.5829146472
6855646, 'CollegeState_Haryana': -0.49861470967037524, 'Gender_f': -1.4753485283428724, 'CollegeS
tate_Delhi': -0.68634096871584249, '12board_mp board': -0.49121424602028058, 'CollegeTier': -0.5001
3343568538682, '12board_ipe': -0.5706961603616344}

SKILL: ElectricalEngg
Train acc: 0.957638425617
Test acc: 0.966
ROC AUC: 0.745250274023

Top 10: {'10board_bihar school examination board patna': 0.45786112658130806, 'CollegeState_Tamil
Nadu': 0.59082449013080307, '12board_bihar school examination board patna': 0.45786112658130806,
'CollegeState_Jharkhand': 0.99711141686217331, '12board_icse': 0.9690185921777007, '12board_nagp
ur': 0.45908368713465031, '10board_secondary state certificate': 0.46879285894396183, '12board_high

er secondary state certificate': 0.46879285894396183, '10board_nagpur': 0.45908368713465031, '10boa
rd_cbse': 0.67391569640984439}

Bottom 10: {'years_since_grad': -0.56154303394898342, 'CollegeState_Haryana': -0.7999064330606050
5, '10board_ssc': -0.54816385453379235, 'CollegeState_Uttarakhand': -0.57354748171113756, 'Gender
_f': -0.70559904084637282, 'CollegeState_Punjab': -0.44025761433058486, '12board_0': -0.7844865794
9488143, '10board_0': -0.6139242844142504, 'CollegeState_Maharashtra': -0.58806644305243649, 'Coll
egeState_Himachal Pradesh': -0.3606194694810812}

SKILL: TelecomEngg
Train acc: 0.908939292862
Test acc: 0.9
ROC AUC: 0.561933333333

Top 10: {'12board_board of technicaleducation ,delhi': 0.76794671155288952, '12board_intermediate sta
te board': 0.77310793353503948, '10board_matriculation': 0.7777189049681057, '12board_bice': 0.7808
5055880738652, '12board_bieap': 0.85691762929894999, '10board_board of secondary education,orissa
': 0.80011870004297292, '12board_chse,orissa': 1.073407280121879, '10board_mp state board': 0.8369
6663100764968, '12board_pu board': 1.2738850219478959, '12board_govt of karnataka department of p
re-university education': 0.80199413181706725}

Bottom 10: {'10board_bseb': -0.51384498963921987, '12board_wbchse': -0.61931853860066655, 'Colle
geState_Andhra Pradesh': -0.52881659866474107, '12board_hbse': -0.49186245776318654, '12board_p
uc': -0.53209964034567792, 'Gender_f': -0.52651725784851788, '12board_mp board': -0.495018944758
21184, '12board_hsc': -0.84665018462815467, '12board_maharashtra board': -0.62366272070663431, '
CollegeState_Himachal Pradesh': -0.62711991377579546}

SKILL: CivilEngg
Train acc: 0.990660440294
Test acc: 0.986
ROC AUC: 0.535786728484

Top 10: {'10board_karnataka state education examination board': 0.79987928566066646, '10board_ssc':
1.1941040089501829, '12board_dept of pre-university education': 0.79987928566066646, 'CollegeState_
Jharkhand': 1.2532321755671474, '12board_intermediate': 0.68742423156717336, '10board_karnataka
secondary education board': 0.88134009170176708, '12board_board of school education uttarakhand': 0.
90568202035068224, '12board_pre-university board': 0.93765495706267865, '10board_board of school
education uttarakhand': 0.90568202035068224, 'CollegeState_Himachal Pradesh': 0.7278258274443030
2}

Bottom 10: {'CollegeState_Haryana': -0.67956984163083567, 'CollegeState_Tamil Nadu': -0.537633625
41951021, 'CollegeState_Orissa': -0.37356522324150437, '10board_cbse': -0.4282384621940038, 'Gen

der_f': -0.52205058688513384, 'CollegeState_Telangana': -0.85396791912216818, 'CollegeCityTier': -0.56610289928661095, 'CollegeState_Rajasthan': -0.45465960593174687, 'CollegeTier': -1.5894774512432277, 'CollegeState_Delhi': -0.59345384729604733}